

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68	A2	(11) International Publication Number: WO 00/58519 (43) International Publication Date: 5 October 2000 (05.10.00)
<p>(21) International Application Number: PCT/US00/08440</p> <p>(22) International Filing Date: 30 March 2000 (30.03.00)</p> <p>(30) Priority Data: 60/127,248 31 March 1999 (31.03.99) US</p> <p>(71) Applicants (<i>for all designated States except US</i>): WHITE-HEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (<i>for US only</i>): ALTSHULER, David [US/US]; 69 Mason Terrace, Brookline, MA 02146 (US). CARGILL, Michele [US/US]; One Warwick Park #3, Cambridge, MA 02140 (US). DALEY, George, Q. [US/US]; 50 Young Road, Weston, MA 02493 (US). IRELAND, James, S. [US/US]; One Warwick Park, #3, Cambridge, MA 02140 (US). LANDER, Eric, S. [US/US]; 151 Bishop Allen Drive, Cambridge, MA 02139 (US). LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US). PATIL, Nila [US/US]; 780 West California Way, Woodside, CA 94602 (US). SKLAR, Pamela [US/US]; 68 Fuller Street, Brookline, MA 02446 (US).</p>		<p>(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).</p> <p>(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>Without international search report and to be republished upon receipt of that report.</i></p>
<p>(54) Title: CHARACTERIZATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN CODING REGIONS OF HUMAN GENES</p> <p>(57) Abstract</p> <p>The invention provides nucleic acid segments of the human genome, particularly nucleic acid segments from the coding region of a gene, including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking or containing these sites are also provided. The nucleic acids, primers and probes are used in applications such as phenotype correlations, forensics, paternity testing, medicine and genetic analysis.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

CHARACTERIZATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN CODING REGIONS OF HUMAN GENES

RELATED APPLICATION

This application claims the benefit of U.S. Provisional Application Serial
5 No. 60/127,248, filed March 31, 1999, the entire teachings of which are incorporated
herein by reference.

GOVERNMENT SUPPORT

The invention was supported, in whole or in part, by grant 5-P50-HG00098-
09 SNP from the National Institutes of Health (NCHGR) and grant 1-R01-
10 HL61774-01 from the National Institutes of Health (NHLBI). The Government has
certain rights in the invention.

BACKGROUND OF THE INVENTION

A major goal in human genetics is to understand the role of common genetic
variants in susceptibility to common diseases (N. Risch and K. Merikangas, *Science*,
15 273:1516 (1996.); E. S. Lander, *Science*, 274:536 (1996); F.S. Collins, *et al.*,
Science, 278:1580 (1997)). This requires assembling an extensive catalogue of
single-nucleotide polymorphisms (SNPs) and performing systematic association
studies for particular diseases.

The human population has relatively limited genetic diversity, reflecting its
20 young age and historically small size (F. J. Ayala *et al.*, *Proc. Natl. Acad. Sci.*,
91:6787 (1994)). Given the restricted nature of the allelic spectrum, some authors
have recently suggested that it should eventually be possible to collect all common
SNPs in the human population and have hypothesized that such common variants
may underlie much of the genetic risk of common disease (N. Risch and K.
25 Merikangas, *Science*, 273:1516 (1996.); E. S. Lander, *Science*, 274:536 (1996); F.S.
Collins, *et al.*, *Science*, 278:1580 (1997)). This is in contrast to the situation for rare

Collins, *et al.*, *Science*, 278:1580 (1997)). This is in contrast to the situation for rare genetic diseases, which are primarily caused by a large number of distinct alleles that are recent, rare and highly penetrant. Important examples of associations to common (>1%) alleles include the ApoE4 allele in Alzheimer's disease, the Factor
5 V^{Leiden} allele in deep-venous thrombosis, and the CCR5-Δ32 in resistance to HIV infection (A. M. Saunders *et al.*, *Neurology*, 43:1467 (1993); R. M. Bertina, *Nature*, 369:64 (1994); M. Dean *et al.*, *Science*, 273:1856 (1996)). The most relevant variants are likely to be those in coding and regulatory regions of genes.

SUMMARY OF THE INVENTION

10 As described herein, the nature of SNPs in the coding regions of human genes has been explored. SNPs were identified in 106 genes relevant to cardiovascular disease, endocrinology and neuropsychiatry, by screening an average of 114 independent alleles using two independent screening methods. To ensure high accuracy, all reported SNPs were confirmed by DNA sequencing. A total of
15 545 SNPs were identified, including 395 coding-regions SNPs (cSNPs) divided roughly equally between those causing synonymous and non-synonymous changes. The cSNPs most likely to influence disease, those that alter the amino acid sequence of the encoded protein, show strikingly different properties: they occur at a lower rate and with lower allele frequencies. This likely reflects selection acting against
20 deleterious alleles during human evolution. The lower allele frequency of cSNPs has important implications for the number of chromosomes that must be sampled to construct a comprehensive catalogue of human cSNPs.

The invention relates to a gene which comprises a single nucleotide polymorphism at a specific location. In a particular embodiment the invention
25 relates to the variant allele of a gene having a single nucleotide polymorphism, which variant allele differs from a reference allele by one nucleotide at the site(s) identified in Figures 5A-5Q. Complements of these nucleic acid segments are also included. The segments can be DNA or RNA, and can be double- or single-stranded. Segments can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30,
30 10-50 or 10-100 bases long. The invention further relates to gene products encoded by genes and oligonucleotides of the invention.

The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a single nucleotide polymorphism or to the complement of the gene. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in Figures 5A-5Q. Optionally, a set of bases occupying a set of the polymorphic sites shown in Figures 5A-5Q is determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the polymorphic site or sites in the individuals tested.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graph showing minor allele frequency by polymorphism type. The percentage of cSNPs having minor allele frequency classified as low (<5%), medium (5-15%) or high (>15%) frequency is displayed for synonymous, non-synonymous and non-coding SNPs.

Figure 2 is a graph showing the distribution of nucleotide diversity. Normalized frequency of variant sites, θ , was calculated for the coding region of each gene. The graph shows the percentage of genes having θ in the indicated range.

Figures 3A and 3B are a table showing a summary of polymorphisms in 106 human genes described herein. Column 1 shows the name of the gene as used in Online Mendelian Inheritance in Man. Column 2 shows the number of coding base pairs screened. Column 3 shows the number of synonymous (or silent) polymorphisms identified. Column 4 shows the number of non-synonymous polymorphisms identified. Column 5 shows the number of non-coding base pairs screened. Column 6 shows the number of non-coding polymorphisms, including those in introns and untranslated regions (UTR), identified.

Figure 4 is a table showing polymorphism rates for different classes of sites.

Figures 5A-5Q are a table showing the specific polymorphisms identified in the genes studied as described herein. Column 1 shows the laboratory

designation for the polymorphism. Column 2 shows the name of the gene as used in Online Mendelian Inheritance in Man. Column 3 shows the reference nucleotide which occupies the polymorphic site in the reference allele. Column 4 shows the variant nucleotide which occupies the polymorphic site in the variant allele.

- 5 Column 5 shows the reference amino acid encoded by the codon which contains the polymorphic site in the reference allele. Column 6 shows the variant or alternate amino acid encoded by the codon which contains the polymorphic site in the variant allele. Column 7 indicates whether the polymorphism is located in the coding or non-coding region of the gene. Column 8 shows the assay number in which the
- 10 polymorphism was assessed. Columns 9 and 10 show the forward and reverse primers, respectively, which were used to identify the polymorphism. Column 11 shows the sequence of the gene used in the assay, with the polymorphic site indicated by brackets and the primers shown in capital letters. Column 12 shows the total number of nucleotides given in Column 11.

15 DETAILED DESCRIPTION OF THE INVENTION

- There is a rich literature concerning nucleotide variation in model systems, particularly in *Drosophila* (E. N. Moriyama and J. R. Powell., *Mol. Biol. Evol.*, 13:261 (1996)), but sequence variation in human genes has been studied only in limited ways. A small number of studies have focused on individual genes (such as
- 20 beta-globin and lipoprotein lipase) in many individuals, and one study examined 49 genes by comparing two independent sequences deposited in public databases (R. M. Harding *et. al.*, *Am. J. Hum. Genet.*, 60:772 (1997); D. A. Nickerson *et. al.*, *Nature Genetics*, 19:233 (1998); W. H. Li and L.A. Sadler, *Genetics* 129:513 (1991)). To perform a more comprehensive survey, as described herein, a collection
- 25 of 106 genes were selected whose protein products play important roles in the cardiovascular, endocrine and neurological systems (Figures 3A-3B and Figures 5A-5Q). Gene sequences were obtained from the Genbank and TIGR databases. Where multiple sequence depositions were available, a consensus sequence was derived. Determination of coding sequence, untranslated regions and
- 30 intronic regions was based on annotation in the public database, although internal

checks were performed to ensure accurate determination of start and stop codons, open reading frames and the like.

The genes were chosen because of their relevance to common, clinically significant diseases, such as coronary artery disease, diabetes, and schizophrenia.

- 5 They encode proteins involved in coagulation, lipid metabolism, energy metabolism, neuroendocrine physiology, neurotransmission and central nervous system development. Variation in these genes was studied in a sample including Caucasians, African-Americans, African Pygmies and Asians, with an average of 114 chromosomes screened for each gene. Of the samples screened, 30 were from
- 10 Caucasian individuals, 14 from Asian, 10 African American and 7 Africans. The average number of individuals successfully screened for each gene was 57, with the precise number successfully screened varying among genes. Cell lines were obtained from Coriell Cell Repository, and DNA prepared according to standard protocols. In addition, 10 of the Caucasian samples used in this study were obtained
- 15 as anonymous blood samples from the Physician's Health Study (gift of Charles Hennekens and J. Michael Gaziano). The sample size provides greater than 65% power to detect alleles with frequency of 1%.

- Overall, the sample of 114 chromosomes was screened for SNPs in a total of 195.4 kb, consisting of 135.8 kb of coding regions and 59.6 kb from adjacent
- 20 non-coding region (untranslated region (UTR) and introns). Sequences were amplified by the polymerase chain reaction (PCR) and screened by two independent methods. The first method involved hybridization of labeled PCR products to variant detector arrays (VDAs) (that is, high density DNA probe arrays containing oligonucleotides specific for the sequences under study) (M. Chee *et. al.*, *Science*,
- 25 274:610 (1996); D. G. Wang *et. al.*, *Science*, 280:1077 (1998)); variant sequences typically give rise to altered hybridization patterns. These chips contained variant detector arrays (VDA) (M. Chee *et al.*, *Science* 274:610 (1996)).

- Using VDAs, candidate SNPs were identified using a combination of three algorithms followed by visual inspection. For each base position and strand queried
- 30 there are four VDA features: one contains the expected base (the reference sequence) in the central position and the other three features contain central substitution bases (in the background of the reference sequence). The base-calling

algorithm looked for positions at which hybridization to a substitution base gives a stronger signal than the reference base. The second algorithm (mutant fraction) examined the reference base and each one of the substitution bases in turn and calculates the fraction of signal present in the non-reference base. The final
5 algorithm (footprint detection) depends upon a loss of signal at the reference positions surrounding a nucleotide substitution. These algorithms are combined to yield a confidence score of "certain" or "likely" for each candidate polymorphism. Two analysts independently scored the data, and candidate polymorphisms found by either observer were included in subsequent confirmation tests. PCR assays
10 spanning each exon were designed using Primer 3.0 release 0.7. PCR was performed according to standard protocols, and assays destined to be hybridized to the same chip design were pooled together. Chip samples were prepared and hybridized as described in D.G. Wang *et al.* (*Science* 280:1077 (1998)), except that pools consisting of about 100 assays contained 5-6 μ g of amplified material. In all,
15 854 assays (average size of 300 bp, covering 106 genes) were amplified from each individual and were hybridized to 12 distinct chip designs. The probe arrays were designed to query only the coding sequence for some genes, while other genes contained the entire mRNA and/or surrounding intron (Figures 3A-3B). The second method involved subjecting PCR products to Denaturing HPLC (dHPLC) (P. J.
20 Oefner and P. A. Underhill, *Am. J. Hum. Genet.*, 57:A266 (1995)) at a critical temperature; heterozygous individuals typically give rise to heteroduplex products with altered denaturation and migration properties.

Sequences were amplified as above except that the final extension in the PCR protocol was followed by denaturation and slow reannealing to allow
25 heteroduplex formation. A total of 6 μ l of each individual PCR product was injected into Wave DNA Fragment Analysis System (Transgenomic). A total of 592 of the VDA assays (covering the 89 genes attempted with this method) were successfully screened by dHPLC. Only assays of >160 base pairs were used for dHPLC, because shorter assays performed unreliably for mutation detection. The
30 dHPLC parameters (percentage of acetonitrile, column temperature) used for each fragment were automatically calculated using a novel predictive algorithm, and dHPLC traces were analyzed using the clustering program ASH v2.0. A scoring

algorithm was developed based upon the similarity score by ASHv2.0 and contour of the elution profile.

Because both screening methods can generate to a significant number of false positives, it was important to confirm every reported SNP. Samples implicated by either method as containing a candidate SNP were thus subjected to fluorescent dideoxy sequencing, either to confirm the presence of the SNP (in the case of the chip) or to identify and confirm the presence of the SNP (in the case of DHPLC). Such confirmation proved essential for eliminating false positives.

Candidate SNPs were either validated (if found by VDAs) or identified (if implicated by DHPLC) by DNA sequencing. For this purpose, sequences were amplified with PCR primers tailed with standard M13 sequencing sites (-21 forward and -28 reverse) and conventional dye-primer sequencing was performed on ABI 377 sequencers. For candidate SNPs discovered by VDAs, one individual was chosen (a candidate homozygous variant, when available, or a candidate heterozygote) and sequencing was performed on one strand to confirm by visual inspection the presence of the SNP at the indicated position. For amplicons found to be polymorphic by DHPLC, two individuals were selected representing each distinct elution pattern observed and were sequenced on both strands to discover the variant base or bases. Sequences were base-called by the Phred program, assembled by the Phrap program, and polymorphism candidates were identified by the PolyPhred program (D. A. Nickerson *et. al.*, *NAR*, 25:2745 (1997)). All results were visually inspected by at least two observers.

The overall false positive rate for VDAs was 45%. The rate was much lower (about 10%) for certain chip designs, synthesis protocols, and for candidate polymorphisms scored as "certain." The false positive rate among fragments displaying an altered elution pattern by DHPLC was similar (40%). The false positive rates reflect the thresholds employed for declaring a candidate SNP, which were chosen to ensure high sensitivity.

A total of 545 SNPs were identified in the 195 kb surveyed, consisting of 150 non-coding SNPs and 395 cSNPs. Results from these studies are shown in the Figures. The complete data are available on the web site http://www.genome.wi.mit.edu/cvar_snps; access to this website can be gained

using the guestname "snp_pilot" and the password "noynek". In the future, access to this website may be available to the public, and thus, no guestname or password may be needed.

To directly determine the false-negative rate of the screen, conventional
 5 DNA sequencing was performed on ten of the genes (THPO, TBAX2R, PTHLH, IGF2, HTR2A, HTR1A, GHR, GABRB1, F10, and CYP11B1) spanning 25.2 kb in twenty individuals. Sequencing was performed on both strands using dye-primer chemistry and sequence traces were interpreted using PolyPhred (D.A. Nickerson *et al.*, *NAR*, 25:2745 (1997)). VDA analysis identified 85% of variants found by direct
 10 sequencing, while DHPLC identified 87% of the variants found by direct sequencing. In regions screened by both VDAs and DHPLC, the combination of the two methods identified 100% of the polymorphisms found by direct sequencing.

Overall, about one-third of individuals were screened with both methods, and one-third were screened with each of the two methods alone. (For some genes, the
 15 non-coding regions were screened only by DHPLC.) It is estimated that the false negative rate over the entire study to be about 15% for regions screened by one method, and negligible for sequences screened by both methods. The total number of true polymorphisms not identified is estimated to be less than 10%.

A SNP survey can be characterized in terms of either K , the observed
 20 number of variant sites, or p , the observed heterozygosity per bp. Because K increases with the number of chromosomes (n) studied and the total sequence length L , it is preferable to use the normalized number of variant sites

$$\hat{\theta} = K / \left(\sum_{i=1}^{n-1} i^{-1} \right) L \text{ which corrects for sample size. Under the neutral}$$

theory of molecular evolution and infinite sites model, θ and π are both estimators
 25 of the population genetic parameter $\theta = 4N\mu$ (Li, *Molecular Evolution*, Sinauer Associates (1997), Canada).

SNPs were found at a similar overall frequency in coding and non-coding regions. SNPs in coding region occurred at a frequency of 1 per 344 bp, corresponding to $\hat{\theta} = 5.47 \times 10^{-4}$ and $\pi = 5.07 \times 10^{-4}$. Interestingly, SNPs were
 30 observed in non-coding DNA at a similar frequency of 1 per 397 bp. The

normalized number of variant sites was $\theta = 4.93 \times 10^{-4}$, and the mean heterozygosity (π) = 5.05×10^{-4} (Figure 4). Calculations of π involve allele frequencies. Polymorphisms identified by DHPLC alone were excluded because we did not sequence all of the samples showing a variant DHPLC pattern and thus could not be certain of allele frequency. The estimates of π were thus based on 411 of 545 polymorphisms. Although the VDAs were designed for polymorphism discovery rather than genotyping, the estimated allele frequencies proved to be quite accurate. Specifically, genotyping assays (employing single-base extension assays) for 25 SNPs yielded allele frequencies that differed by an average of only 2% from those estimated on the basis of genotypes inferred from the VDA. For both classes, the similar values for θ and π is consistent with a population evolving according to neutral expectations.

The 395 cSNPs were roughly equally divided between synonymous (203 cSNPs) and non-synonymous (192 cSNPs) changes. Since approximately two-thirds of random mutations would alter an amino acid, the fact that non-synonymous cSNPs comprise slightly less than half of the cSNPs implies strong selection against amino-acid altering changes. To address this issue more directly, the nucleotide diversity was examined at four-fold degenerate sites, two-fold degenerate sites, and non-degenerate sites. Changes at four-fold degenerate sites produce only synonymous changes, while those at non-degenerate sites are always non-synonymous. Nucleotide diversity (θ) was 9.64×10^{-4} at four-fold degenerate sites, 6.85×10^{-4} at two-fold degenerate sites, and 3.70×10^{-4} at non-degenerate sites. Assuming that mutations occur at an equal rate at both classes of sites, non-synonymous variants survive to be detected in such a survey at only 38% of the rate of synonymous changes.

The force of selection is also evident in comparing non-synonymous cSNPs causing a non-conservative amino acid alteration with those causing a conservative amino-acid change. Conservative and non-conservative amino acid substitutions were defined for this analysis according to the BLOSUM62 matrix, used in sequence comparison (S. Henikoff and J. G. Henikoff, *PNAS*, 89:10915 (1992)). Conservative changes were those having a positive or neutral sign in the matrix, while non-conservative changes were those having a negative value. Non-conservative

cSNPs represent only 36% of the non-synonymous cSNPs, whereas randomly distributed mutations would be expected to produce a higher proportion (52%) of non-conservative changes. The proportion of non-synonymous SNPs expected to cause a non-conservative amino acid substitution was determined based on the actual codon usage in the 106 genes studied, the known frequencies of transitions and transversions, and the definition of non-conservative changes employed in the BLOSUM62 matrix. This implies that non-conservative cSNPs survive to be detected in such a survey at only about half of the rate of conservative, non-synonymous cSNPs.

The various types of SNPs differ not only in the rate of their occurrence, but also in the frequency of their minor alleles. This can be seen in several ways. When SNPs are classified according to whether the frequency of the minor allele was high ($\geq 15\%$), intermediate (5-15%) or low ($\leq 5\%$), it is clear that the non-synonymous cSNPs were enriched in low frequency alleles compared to the rest of the collection (Figure 1). The distribution of non-synonymous allele frequencies was significantly different than that of synonymous changes ($p=0.02$, Kolmogorov-Smirnov test). Indeed, more than half (58%) of non-synonymous cSNPs were found at a frequency below 5%, with this effect evident for both conservative and non-conservative substitutions.

The effect of selection can also be inferred by considering the average frequency of the minor allele: it is 8% for non-conservative cSNPs, 11% for conservative but non-synonymous cSNPs, and 14% for both synonymous cSNPs and non-coding SNPs. In addition, the lower allele frequency of non-synonymous cSNPs is reflected in the fact that the heterozygosity π is lower than the normalized rate of variant sites $\hat{\theta}$ for this class of SNPs (Figure 4). This divergence is in the direction predicted by the action of purifying selection, although it falls short of statistical significance. Tajima's D was non-significant. (F. Tajima, *Genetics*, 123:545 (1989).

The distribution of SNPs among the 106 genes was explored, with an eye toward detecting differential effects of selection among genes. The number of cSNPs per gene ranged from 37 for Factor V to 0 for thirteen of the genes, and the normalized rate, $\hat{\theta}$, similarly showed considerable variation (Figure 2). The

observed variation in nucleotide diversity is similar in magnitude to that observed for *Drosophila* (E. N. Moriyama and J. R. Powell., *Mol. Biol. Evol.*, 13:261 (1996)). Variation among genes could be due to many factors (D. J. Begun and C. F. Aquadro, *Nature*, 356:519 (1993); Nachman *et. al.*, *Genetics*, 150:1133 (1998)).

- 5 The fact that non-synonymous cSNPs show a somewhat wider variation than synonymous cSNPs (the coefficient of variation is 20% larger for the former class) is consistent with differences in selective constraints among loci, but the difference falls well below statistical significance. A variety of population genetic tests are available for testing selection at individual loci (M. L. Wayne and K. L. Simonson, 10 *Trends and Ecology and Evolution*, 13:236 (1998)).

- The age of a SNP allele has important implications for its use in human genetic studies. Recently-occurring SNP alleles are more likely to show extensive linkage disequilibrium (retention of the ancestral haplotype on which they arose) as compared to older SNPs. Such linkage disequilibrium can provide a powerful tool 15 in identifying disease genes (E. S. Lander, N.J. Schork, *Science*, 265:2037 (1994)). Although the precise age of the SNPs could not be assessed from these studies, characterization of which allele preceded human speciation and which arose thereafter was sought. To determine the ancestral human allele, each corresponding gene was sequenced from the common chimpanzee (*P. troglodytes*). Each assay 20 used in the human survey was amplified from a single chimpanzee (DNA gift of Kristin Ardlie) and subjected to dye-primer sequencing on both strands. A single chimpanzee sample will accurately reveal the ancestral allele except in cases where the site has mutated and fixed during the chimpanzee evolution or is polymorphic in the chimpanzee population and happened to be homozygous for the non-ancestral 25 allele. These two cases are quite rare (probably less than 2%) and thus have been neglected for the purpose of estimating overall rates. A human allele was considered to be ancestral if it was present in the homozygous state in the chimpanzee sample. A total of 136 kb of chimpanzee sequence was obtained, revealing an inter-species divergence of 0.6% in the regions studied.

- 30 An elegant result in theoretical population genetics predicts that the probability that a neutral allele represents the ancestral state should be equal to its frequency in the population (G. A. Watterson and H. A. Guess, *Theoretical*

Population Biology, 11:141 (1977)). The minor allele should thus represent the ancestral state in a predictable proportion of cases. The ancestral allele and minor allele frequency was determined for 267 of the reported SNPs. For 3 of the 267 SNPs, the chimpanzee was homozygous for a third allele differing from both of the
5 current human alleles. This is consistent with the overall 0.6% nucleotide sequence divergences seen between human and chimpanzee. Among polymorphisms with a minor allele frequency below 10%, the average allele frequency was 3% and the proportion that was ancestral was 7% (11/158) of cases. Among polymorphisms with minor alleles exceeding 10%, the mean frequency was 28% and the proportion
10 that were ancestral was 32% (35/109). These results thus agree remarkably well with the theoretical prediction, providing the first reported test of this prediction in humans. It therefore follows that the minor SNP allele need not be the younger allele; this has implications for linkage disequilibrium mapping.

The distribution of SNPs among Caucasian, African-American, African and
15 Asian samples was also examined. Although the vast majority of SNPs were seen in multiple groups, there was a statistically significant excess of SNPs that were seen in only one of the sub-groups. The probability that a SNP occurring $k > 1$ times in an overall sample of n individuals would be found entirely within a given subset of m individuals is $B(n,k)/B(m,k)$, where $B(x,y)$ is the binomial coefficient $x!/(x-y)!y!$. In
20 this fashion, the probability that each individual SNP would be confined to a particular ethnic subgroup within the sample was calculated and these probabilities were summed to obtain the number of SNPs expected to be confined to the group within the sample. The fact that a SNP is found only within one group in the sample does not necessarily imply that it is private to that group within the general
25 population, owing to the small sample size, but it can be used as an indication of substructure. The number of SNPs with $k > 1$ confined to the, African-Americans, African Pygmies, Caucasians, and Asians was 17, 17, 12, and 9, as compared to expectations of 3.02, 1.34, 8.62, and 1.81. Not surprisingly, the greatest excess was seen for SNPs found in the African-American and African samples. The presence of
30 population substructure implies that construction of a comprehensive SNP database should employ a diverse set of DNA samples.

The results of this survey provide a fundamental description of sequence variation in the coding regions of human genes. These data indicate that two copies of a gene chosen from the human population will differ by roughly one base in 2 kb, corresponding to somewhat less than one heterozygous base within the coding region of a typical gene. In general, there are only a handful of such cSNPs per gene that exhibit allele frequencies of at least a few percent. Accounting for both the different rate and frequency of non-synonymous SNPs, only about 40% of these observed changes will alter the encoded amino acid. The action of purifying selection during human evolution is evident from the comparatively lower rate of non-synonymous cSNPs, and especially of those that create a non-conservative change. It is clear that non-synonymous cSNPs not only occur less often, but also have lower minor allele frequencies: 60% of non-synonymous cSNPs, the class likely to have the most dramatic effects on proteins, display a minor allele frequency below 5%.

The relative rarity of cSNPs has important implications for efforts to produce large catalogues of human variants. It has been proposed that most human SNPs could be found by performing shotgun sequencing on a handful of individuals (J. L. Weber and E. W. Myers, *Genome Research*, 7:401 (1997); J. C. Venter *et al.*, *Science*, 280:1540 (1998)). Although such a project will surely identify many SNPs, results described herein suggest that the small sample size will likely fail to identify the vast majority of cSNPs likely to have the most important biological consequences, owing to their lower average allele frequencies. A comprehensive collection of the common, non-conservative cSNPs may require surveying 50-100 chromosomes. Because coding sequence represents only about 3% of the genome, it may prove inefficient to obtain such deep coverage of cSNPs by shotgun sequencing of genomic DNA. Instead, it may be more efficient to perform shotgun sequencing on cDNA libraries from multiple individuals or to amplify genes from multiple individuals, as done here.

Interestingly, a similar rate of polymorphism in coding and non-coding DNA was found. Furthermore, the observed rate of nucleotide diversity at four-fold degenerate sites was nearly twice that in adjacent non-coding regions, and over twice that at non-degenerate sites (Figure 4). Similar results have been reported for

Drosophila (E. N. Moriyama and J. R. Powell., *Mol. Biol. Evol.*, 13:261 (1996)) and for a smaller human data set by Li and Sadler (R. M. Harding *et. al.*, *Am. J. Hum. Genet.*, 60:772 (1997); D. A. Nickerson *et. al.*, *Nature Genetics*, 19:233 (1998); W. H. Li and L.A. Sadler, *Genetics* 129:513 (1991)), who observed over three times
5 the nucleotide diversity at four-fold degenerate sites ($\theta = 11 \times 10^{-4}$), as compared to that in both untranslated regions and non-degenerate sites ($\theta = 3 \times 10^{-4}$). These observations suggest that non-coding DNA adjacent to coding regions may be functionally constrained to a surprising degree.

SNPs can be used to search for genes underlying complex traits in two
10 distinct ways: linkage disequilibrium (LD) studies and association studies (E. S. Lander, N.J. Schork, *Science*, 265:2037 (1994)). Genome-wide LD studies involve using a dense collection of SNPs as markers to search for an ancestral haplotype carrying a disease-susceptibility allele. Such studies cannot be undertaken without the availability of an extremely dense SNP map and their potential for success
15 depends sensitively on many population genetic assumptions. Association studies are more straightforward because they directly test the hypothesis that a specific SNP increases disease risk. They make few assumptions, and require only the availability of a suitable database of appropriate SNPs. In the near term, focusing on cSNPs is likely to be most productive inasmuch as the class is easily recognized (in
20 contrast to regulatory polymorphisms) and is likely to contain a significant proportion of the disease-susceptibility alleles.

The present invention relates to a gene which comprises a single nucleotide polymorphism (SNP) at a specific location. The gene which includes the SNP has at least two alleles, referred to herein as the reference allele and the variant allele. The
25 reference allele (prototypical or wild type allele) has been designated arbitrarily and typically corresponds to the nucleotide sequence of the gene which has been deposited with GenBank or TIGR under a given Accession number. The variant allele differs from the reference allele by one at least one nucleotide at the site(s) identified in Figures 5A-5Q. The present invention also relates to variant
30 alleles of the described genes and to complements of the variant alleles. The invention further relates to portions of the variant alleles and portions of complements of the variant alleles which comprise (encompass) the site of the SNP

and are at least 5 nucleotides in length. Portions can be, for example, 5-10, 5-15, 10-20, 5-25, 10-30, 10-50 or 10-100 bases long. For example, a portion of a variant allele which is 5 nucleotides in length includes the single nucleotide polymorphism (the nucleotide which differs from the reference allele at that site) and four
5 additional nucleotides which flank the site in the variant allele. These nucleotides can be on one or both sides of the polymorphism. Polymorphisms which are the subject of this invention are defined in Figures 5A-5Q in QQQQQQ with respect to the reference sequence deposited in GenBank under the Accession number indicated. For example, the invention relates to a portion of a gene (e.g., AADC) having a
10 partial nucleotide sequence as shown in Figures 5A-5Q in QQQQQQ comprising a single nucleotide polymorphism at a specific position. The reference nucleotide for AADC is shown in column 3 and the variant nucleotide is shown in column 4 of Figures 5A-5Q in QQQQQQ. The nucleotide sequences of the invention can be double- or single-stranded.

15 The invention further provides allele-specific oligonucleotides that hybridize to a gene comprising a single nucleotide polymorphism or to the complement of the gene. These oligonucleotides can be probes or primers.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the
20 polymorphic sites shown in Figures 5A-5Q in QQQQQQ. Optionally, a set of bases occupying a set of the polymorphic sites shown in Figures 5A-5Q in QQQQQQ is determined. This type of analysis can be performed on a number of individuals, who are tested for the presence of a disease phenotype. The presence or absence of disease phenotype is then correlated with a base or set of bases present at the
25 polymorphic site or sites in the individuals tested.

An oligonucleotide of this invention can be DNA or RNA, and single- or double-stranded. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred oligonucleotides of the invention include segments of DNA, or their complements, which include any one of the
30 polymorphic sites shown in Figures 5A-5Q in QQQQQQ. The segments can be between 5 and 250 bases, and, in specific embodiments, are between 5-10, 5-20, 10-20, 10-50, 20-50 or 10-100 bases. The polymorphic site can occur within any

position of the segment. The segments can be from any of the allelic forms of DNA shown in Figures 5A-5QQQQQQQ.

As used herein, the terms "nucleotide" and "nucleic acid" are intended to be equivalent. The terms "nucleotide sequence", "nucleic acid sequence" "nucleic acid
5 molecule" and "segment" are intended to be equivalent.

Hybridization probes are oligonucleotides which bind in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen *et al.*, *Science* 254, 1497-1500 (1991). Probes can be any length suitable for specific hybridization to the target nucleic acid
10 sequence. The most appropriate length of the probe may vary depending upon the hybridization method in which it is being used; for example, particular lengths may be more appropriate for use in microfabricated arrays, while other lengths may be more suitable for use in classical hybridization methods. Suitable probes and primers can range from about 5 nucleotides to about 30 nucleotides in length. For
15 example, probes and primers can be 5, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 25, 26, 28 or 30 nucleotides in length. The probe or primer preferably contains at least one polymorphic site occupied by any of the possible variant nucleotides. The nucleotide sequence can correspond to the coding sequence of the allele or to the complement of the coding sequence of the allele.

As used herein, the term "primer" refers to a single-stranded oligonucleotide which acts as a point of initiation of template-directed DNA synthesis under appropriate conditions (*e.g.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The
25 appropriate length of a primer depends on the intended use of the primer, but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template, but must be sufficiently complementary to hybridize with a template. The term primer site refers
30 to the area of the target DNA to which a primer hybridizes. The term primer pair refers to a set of primers including a 5' (upstream) primer that hybridizes with the 5'

end of the DNA sequence to be amplified and a 3' (downstream) primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

As used herein, linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome. It can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

As used herein, polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic or biallelic polymorphism has two forms. A triallelic polymorphism has three forms.

By altering amino acid sequence, SNPs may alter the function of the encoded proteins. The discovery of the SNP facilitates biochemical analysis of the variants and the development of assays to characterize the variants and to screen for pharmaceutical that would interact directly with on or another form of the protein. SNPs (including silent SNPs) may also alter the regulation of the gene at the transcriptional or post-transcriptional level. SNPs (including silent SNPs) also enable the development of specific DNA, RNA, or protein-based diagnostics that detect the presence or absence of the polymorphism in particular conditions.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site

is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele. Typically the polymorphic site is occupied by a base other than the reference base. For example, where the reference allele contains the base "T" at the polymorphic site, the altered allele can contain a "C", "G" or "A" at the polymorphic site.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C, or equivalent conditions, are suitable for allele-specific probe hybridizations. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleotide sequence and the primer or probe used.

The term "isolated" is used herein to indicate that the material in question exists in a physical milieu distinct from that in which it occurs in nature. For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstance, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present.

I. Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than
5 pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

10 Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et*
15 *al.*, *Nucleic Acids Res.* 19, 4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren *et al.*, *Science* 241, 1077
20 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA*, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and
25 double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis of target DNA for detecting polymorphisms. The first type of analysis, sometimes referred to as de novo
30 characterization, is carried out to identify polymorphic sites not previously characterized (i.e., to identify new polymorphisms). This analysis compares target

sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such alleles/haplotypes in the population can be determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The de novo identification of polymorphisms of the invention is described in the Examples section. The second type of analysis determines which form(s) of a characterized (known) polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki *et al.*, *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15-mer at the 7 position; in a 16-mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some examples of which are described in WO 95/11995. One form of such arrays is described in the Examples section in connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant form of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples, except that the probes exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particularly useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (e.g., two or more mutations within 9 to 21 bases).

3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, e.g., WO 93/22456).

4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook *et al.*, *Molecular Cloning, A Laboratory*

- 5 *Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can
10 be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

6. Single-Strand Conformation Polymorphism Analysis

- 15 Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single
20 stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence differences between alleles of target sequences.

II. Methods of Use

- 25 After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a set of polymorphic forms that distinguishes the individual. See generally National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard *et al.*, National Academy Press, DC, 1996). The more sites that are analyzed, the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred polymorphisms for use in forensics are biallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

$p(ID)$ is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In biallelic loci, four genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies x and y , the probability of each genotype in a diploid organism is (see WO 95/12607):

$$\text{Homozygote: } p(AA) = x^2$$

$$\text{Homozygote: } p(BB) = y^2 = (1-x)^2$$

Single Heterozygote: $p(AB) = p(BA) = xy = x(1-x)$

Both Heterozygotes: $p(AB+BA) = 2xy = 2x(1-x)$

- The probability of identity at one locus (i.e., the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation:

$$p(ID) = (x^2)^2 + (2xy)^2 + (y^2)^2.$$

- These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity $p(ID)$ for a 3-allele system where the alleles have the frequencies in the population of x , y and z , respectively, is equal to the sum of the squares of the genotype frequencies:

$$p(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + y^4 + z^4$$

In a locus of n alleles, the appropriate binomial expansion is used to calculate $p(ID)$ and $p(exc)$.

- The cumulative probability of identity (cum $p(ID)$) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus.

$$\text{cum } p(ID) = p(ID1)p(ID2)p(ID3).... p(IDn)$$

The cumulative probability of non-identity for n loci (i.e. the probability that two random individuals will be different at 1 or more loci) is given by the equation:

$$\text{cum } p(\text{nonID}) = 1 - \text{cum } p(ID).$$

- If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (e.g., one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

- The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the set of polymorphisms of the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

$$p(\text{exc}) = xy(1-xy)$$

where x and y are the population frequencies of alleles A and B of a biallelic polymorphic site.

(At a triallelic site $p(\text{exc}) = xy(1-xy) + yz(1-yz) + xz(1-xz) + 3xyz(1-xyz)$), where x, y and z are the respective population frequencies of alleles A, B and C).

The probability of non-exclusion is

$$p(\text{non-exc}) = 1 - p(\text{exc})$$

The cumulative probability of non-exclusion (representing the value obtained when n loci are used) is thus:

$$\text{cum } p(\text{non-exc}) = p(\text{non-exc1})p(\text{non-exc2})p(\text{non-exc3})\dots p(\text{non-excn})$$

The cumulative probability of exclusion for n loci (representing the probability that a random male will be excluded)

$$\text{cum } p(\text{exc}) = 1 - \text{cum } p(\text{non-exc}).$$

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his/her father.

C. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending

on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single
5 polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped
10 genetic components. Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of diseases which can be treated or diagnosed as described herein include, but are not limited to,
15 bradyarrhythmias, tachyarrhythmias, heart failure, such as congestive heart failure, congenital heart disease, rheumatic fever, valvular heart disease, cardiomyopathies, myocarditides, pericardial diseases, cardiac tumors, cardiac manifestations of systemic diseases, and traumatic cardiac injury. Other disorders include atherosclerosis, acute myocardial infarction, ischemic heart disease, hypertensive
20 vascular disease, disorders of the aorta, vascular diseases of the extremities, vessel wall disorders, such as various forms of thrombocytopenia, von Willebrand's disease and drug-induced platelet dysfunction, and homeostatic disorders relating to vessel disease and associated bleeding. Also suitable are thrombotic thrombocytopenic purpura, hemolytic-uremic syndrome, Henoch-Schönlein purpura, capillary fragility,
25 vascular purpura, metabolic and inflammatory disorders, such as those induced by rickettsiae and certain drugs, such as sulfonamides, aortic aneurysm, aortic dissection, aortic occlusion, aortitis, atherosclerosis, coronary artery disease, angina, myocardial infarction, thrombosis, hemostatic and coagulation disorders, hypertension and hypotension. Other disorders include transplant accelerated
30 vascular restenosis following balloon angioplasty, Raynaud's disease and acrocyanosis.

Additional disorders include, but are not limited to, disorders of neurodegeneration characterized by astrocyte hypertrophy including gliosis, Pick's disease, aceroplasminemia, portal-systemic encephalopathy, frontal lobe dementia and inherited and acquired ataxias, neurodegenerative diseases of other etiology including progressive supranuclear palsy, primary progressive aphasia, cortical basal degeneration, Alzheimer's disease, Huntington's disease, and Parkinson's disease, retinitis pigmentosa and amyotrophic lateral sclerosis. Other disorders include epilepsy, stroke, defects of neural migration and differentiation, including Miller-Dieker lissencephaly syndrome, and cancer of the brain including astrocytomas and gliomas, as well as psychological disorders such as schizophrenia.

Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

The correlation of one or more polymorphisms with phenotypic traits can be facilitated by knowledge of the gene product of the wild type (reference) gene. The genes in which cSNPs of the present invention have been identified are genes which have been previously sequenced and characterized in one of their allelic forms.

Correlation is performed for a population of individuals who have been tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a χ -squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a further example, it might be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo in vitro fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that this treatment regime should be followed.

For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz *et al.*, US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \dots \beta_{17} + PE_n + a_n + e_p$$

where Y_{ijkpn} is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record; μ is an overall mean; YS_i is the effect common to all cows calving in year-season; X_k is the effect common to cows in either the high or average selection line; β_1 to β_{17} are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms; PE_n is permanent environmental effect common to all records of cow n ; a_n is effect of animal n and is

composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and e_p is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used
5 as parents for breeding the next generation of the herd.

D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus
10 associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. See Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 83,
15 7353-7357 (1986); Lander *et al.*, *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller *et al.*, *Cell* 51, 319-337 (1987); Lander *et al.*, *Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. See Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992).

20 Linkage studies are typically performed on members of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic markers co-segregate with a phenotypic trait. See, e.g., Kerem *et al.*, *Science* 245, 1073-1080
25 (1989); Monaco *et al.*, *Nature* 316, 842 (1985); Yamoka *et al.*, *Neurology* 40, 222-226 (1990); Rossiter *et al.*, *FASEB Journal* 5, 21-27 (1991).

Linkage is analyzed by calculation of LOD (log of the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker and a genetic locus when the two are located at a recombination fraction θ , versus
30 the situation in which the two are not linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company,

Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of likelihood ratios are calculated at various recombination fractions (θ), ranging from $\theta = 0.0$ (coincident loci) to $\theta = 0.50$ (unlinked). Thus, the likelihood at a given value of θ is: probability of data if loci linked at θ to probability of data if loci
5 unlinked. The computed likelihoods are usually expressed as the \log_{10} of this ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different families to be combined by simple addition. Computer
10 programs are available for the calculation of lod scores for differing values of θ (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. See Smith *et al.*, *Mathematical tables for research workers in human genetics* (Churchill, London, 1961); Smith, *Ann. Hum. Genet.* 32, 127-150
15 (1968). The value of θ at which the lod score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of θ) than the possibility that the two loci are unlinked. By convention, a combined lod score of
20 +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the
25 remaining non-excluded chromosomal locations.

III. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Figures 5A-5Q, column 11, in which the polymorphic position is
30 occupied by one of the alternative bases for that position. Some nucleic acids encode full-length variant forms of proteins. Similarly, variant proteins have the

prototypical amino acid sequences encoded by nucleic acid sequences shown in Figures 5A-5QQQQQQQ, column 11, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in Figures 5A-5QQQQQQQ. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in Figures 5A-5QQQQQQQ.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles*

and Practice, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

- 5 The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. See Hogan *et al.*,
10 "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. See Capecchi, *Science* 244, 1288-1292 (1989). The transgene is then introduced into an embryonic stem cell, where it undergoes homologous
15 recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

- In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the
20 interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

- 25 Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*,
30 Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and

lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

IV. Kits

5 The invention further provides kits comprising at least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific
10 oligonucleotide probes for detecting at least 10, 100 or all of the polymorphisms shown in Figures 5A-5Q. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and
15 the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

From the foregoing, it is apparent that the invention includes a number of general uses that can be expressed concisely as follows. The invention provides for the use of any of the nucleic acid segments described above in the diagnosis or
20 monitoring of diseases, such as coronary artery disease, diabetes, coagulation disorders, lipid metabolism disorders, energy metabolism disorders, diseases of the blood, blood vessels and cardiovascular system, and infection by microorganisms, as well as psychological disorders (e.g., bipolar disorder, psychiophrenia). The invention further provides for the use of any of the nucleic acid segments in the
25 manufacture of a medicament for the treatment or prophylaxis of such diseases. The invention further provides for the use of any of the DNA segments as a pharmaceutical.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled
30 in the art that various changes in form and details may be made therein without

departing from the spirit and scope of the invention as defined by the appended claims.

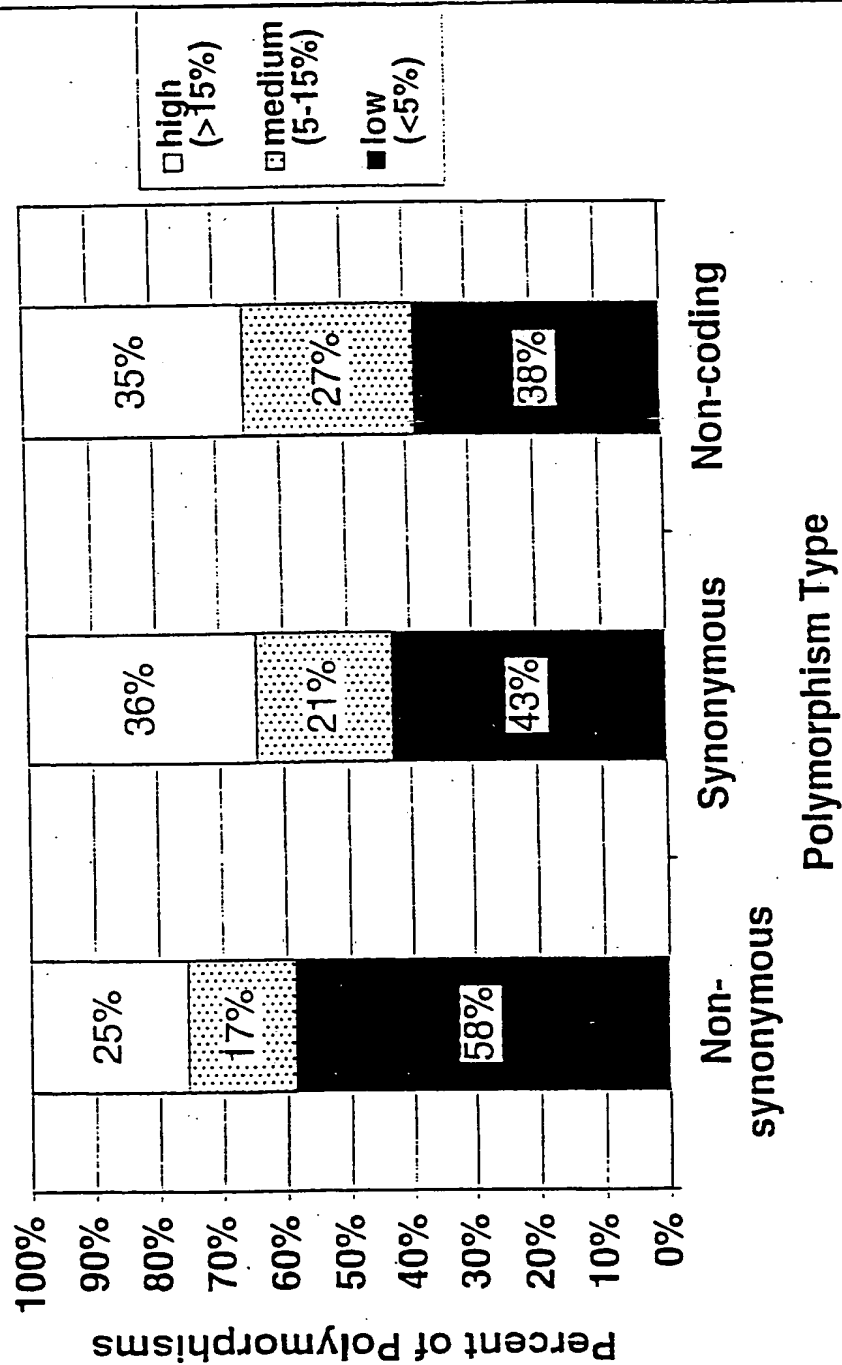
CLAIMS

What is claimed is:

1. A nucleic acid molecule selected from the group consisting of the genes listed in Figures 5A-5Q, wherein said nucleic acid molecule is at least 5 nucleotides in length and comprises a polymorphic site identified in Figures 5A-5Q, wherein a nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.
2. A nucleic acid molecule according to Claim 1, wherein said nucleic acid molecule is at least 10 nucleotides in length.
3. A nucleic acid molecule according to Claim 1, wherein said nucleic acid molecule is at least 20 nucleotides in length.
4. A nucleic acid molecule according to Claim 1, wherein the nucleotide at the polymorphic site is the variant nucleotide for the gene listed in Figures 5A-5Q.
5. An allele-specific oligonucleotide that hybridizes to a portion of a gene selected from the group consisting of the genes listed in Figures 5A-5Q, wherein said portion is at least 5 nucleotides in length and comprises a polymorphic site identified in Figures 5A-5Q, wherein a nucleotide at the polymorphic site is different from a nucleotide at the polymorphic site in a corresponding reference allele.
6. An allele-specific oligonucleotide according to Claim 5 that is a probe.
7. An allele-specific oligonucleotide according to Claim 5, wherein a central position of the probe aligns with the polymorphic site of the portion.

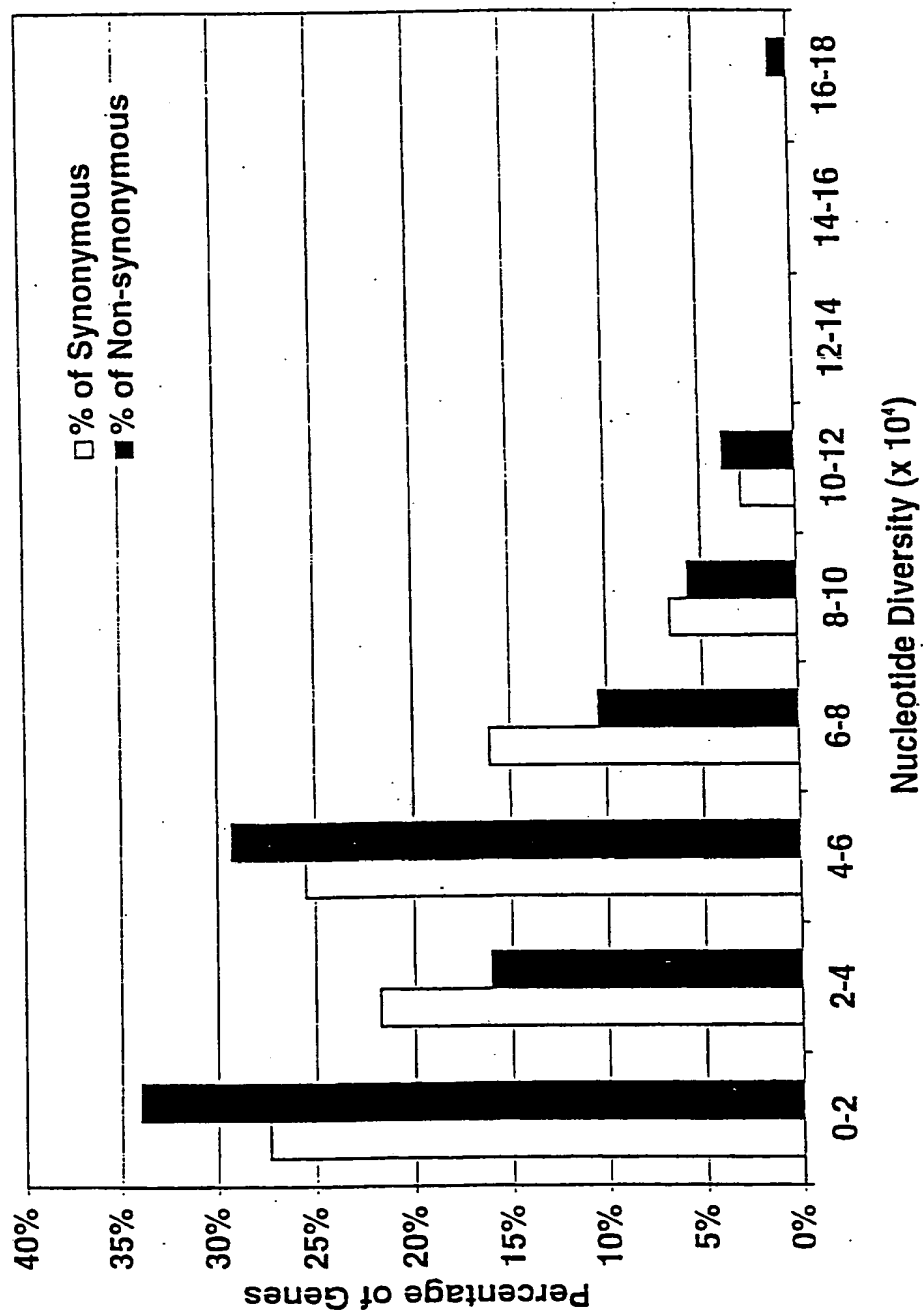
8. An allele-specific oligonucleotide according to Claim 5 that is a primer.
9. An allele-specific oligonucleotide according to Claim 8, wherein the 3' end of the primer aligns with the polymorphic site of the portion.
10. An isolated gene product encoded by a nucleic acid molecule according to Claim 1.
- 5 11. A method of analyzing a nucleic acid sample, comprising obtaining the nucleic acid from an individual sample; and determining a base occupying any one of the polymorphic sites shown in Figures 5A-5Q. Q. Q. Q. Q. Q. Q. Q. Q.
- 10 12. A method according to Claim 11, wherein the nucleic acid sample is obtained from a plurality of individuals, and a base occupying one of the polymorphic positions is determined in each of the individuals, and the method further comprising testing each individual for the presence of a disease phenotype, and correlating the presence of the disease phenotype with the base.

FIG. 1



2/178

FIG. 2



Gene	coding bp screened	No. Synonymous polymorphisms	No. Non- synonymous polymorphisms	Non-coding bp screened	No. Non-coding polymorphisms
AADC	1229	0	2	311	0
ADORA2	332	0	1	75	0
AHC	1413	0	0	63	1
ANX3	929	2	4	725	6
APOD	570	1	3	383	1
AR	2759	3	1	300	0
AT3	1357	3	0	121	0
BDNF	744	0	1	212	0
CD36	1209	1	1	252	0
CETP	1397	4	4	299	0
CGA	349	1	0	235	0
CLanalog	1461	3	2	12	0
CNTF	603	0	1	154	0
COMT	783	2	1	241	1
CRH	51	0	0	745	3
CYP11A	1556	1	1	547	0
CYP11B1	1410	7	7	496	9
CYP11B2	1512	7	8	906	4
CYP17	1395	3	0	36	0
CYP21	1488	5	11	542	7
DBH	1266	0	2	49	0
DRD1	1341	1	0	81	0
DRD2	1032	2	0	1379	3
DRD3	719	0	1	145	0
DRD5	1408	2	1	34	0
F10	1369	3	2	416	1
F11	1878	7	4	1312	2
F13A1	2199	3	6	948	4
F13B	1952	4	6	2339	4
F2	1740	3	2	292	0
F2R	1202	2	1	13	0
F3	875	0	1	92	0
F5	6564	13	16	1542	8
F7	1262	4	2	1209	2
F9	1364	0	1	1062	2
FGA	1935	2	2	490	0
FGB	1476	7	3	1057	0
FGG	1252	0	2	1392	2
FSH	355	1	1	44	0
FSHR	1683	1	3	0	0
GABRB1	1425	5	0	804	2
GAP43	675	1	1	79	0
GH1	644	0	1	426	5
GHR	1765	1	6	391	1
GNRHR	237	0	1	513	0
GP1BA	1881	2	2	48	0
GP1BB	1238	0	0	73	0
GP5	1683	0	0	52	0
GP9	534	1	0	143	0
GRF	224	0	0	239	0
GRIN1	1681	1	0	553	0
GRL	2334	4	3	4028	5
HCF2	1500	3	3	64	1

FIG. 3A

4/178

Gene	coding bp screened	No. Synonymous polymorphisms	No. Non- synonymous polymorphisms	Non-coding bp screened	No. Non-coding polymorphisms
HMGCR	1724	0	1	12	1
HSD3B1	1122	3	2	653	1
HSD3B2	1122	1	1	723	2
HTR1A	1272	1	0	1189	1
HTR1D	1134	1	1	46	0
HTR1DB	1173	2	0	85	1
HTR1E	1098	1	1	70	0
HTR1EL	1101	1	0	46	0
HTR2A	1398	2	3	1709	9
HTR2C	1245	0	1	138	0
HTR5A	1062	2	0	34	0
HTR6	437	1	0	34	0
HTR7	1279	0	0	138	0
IGF1	630	0	0	7250	8
IGF2	546	0	0	610	1
ITGA2B	2833	4	3	707	0
ITGB3	2131	4	3	163	0
KLK2	297	0	1	279	2
LCAT	1289	1	2	90	0
LDLR	2101	7	3	38	0
LIPC	1471	4	3	754	4
LPL	409	1	1	48	0
MAOA	1032	1	0	69	0
MAOB	980	1	0	135	0
MPL	1748	1	2	903	1
NGFB	726	1	1	1186	5
NOS1	127	0	0	56	0
NT3	774	1	0	150	0
NTRK1	1961	5	2	1106	0
PACE	1500	2	0	1095	4
PAI1	1171	1	2	911	1
PAI2	1248	5	4	915	5
PC1	1881	1	3	456	1
PCI	1221	5	5	576	4
POMC	132	0	0	520	0
PRL	633	1	1	180	1
PROC	1334	3	0	114	0
PROS1	1868	1	0	557	0
PTAFR	1029	0	2	13	0
PTH	348	1	0	230	2
PTHLH	634	0	0	2342	13
SELP	2096	5	8	14	0
SHBG	1209	1	3	494	1
SLC6A1	1388	2	0	547	2
SLC6A3	1496	6	1	205	0
SLC6A4	1623	1	2	824	1
TBXA2R	1006	1	0	12	0
TBXAS1	1605	1	6	1411	1
TFPI	806	0	1	139	0
TH	965	1	1	104	0
THBD	1728	0	0	26	0
THPO	1049	0	0	632	2
VLDLR	2391	3	1	850	2
ALL GENES	135823	203	192	59552	150

FIG. 3B

5/178

Polymorphism rates for different classes of sites. Nucleotide diversity and heterozygosity (π) are expressed $\times 10^4$.

Polymorphism Type	bp screened	No. polys	Adjusted for frequency of sites*			
			$\hat{\theta}$	π	Frequency (SNP/bp)	$\hat{\theta}$
Non-coding	59,552	150	4.93 \pm 1.24	5.05 \pm 2.40		
Coding	135,823	395	5.47 \pm 1.32	5.07 \pm 2.40		
synonymous		203	2.81 \pm 0.68	2.98 \pm 1.42	1/191	9.84 \pm 2.38
non-synonymous		192	2.66 \pm 0.64	2.06 \pm 0.98	1/504	3.73 \pm 0.90
conservative		122	1.69 \pm 0.41	1.44 \pm 0.68	1/389	4.94 \pm 1.19
non-conservative		70	0.97 \pm 0.23	0.63 \pm 0.30	1/705	2.61 \pm 0.63
four-fold degenerate sites	21,645	111	9.64 \pm 2.32	9.26 \pm 4.40		
two-fold degenerate sites	34,294	125	6.85 \pm 1.65	5.33 \pm 2.53		
non-degenerate sites	79,659	157	3.70 \pm 0.89	2.52 \pm 1.19		
Total	195,375	545	5.31 \pm 1.28	5.01 \pm 2.38		

* The number of synonymous sites was calculated as the sum of four-fold degenerate sites and half the number of two-fold degenerate sites; the number of non-synonymous sites is the sum of the non-degenerate sites and half the two-fold degenerate sites. The number of conservative and non-conservative sites is estimated as the proportion of non-synonymous sites at which a nucleotide substitution would create a conservative or non-conservative substitution, calculated as in footnote 21.

FIG. 4

9/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer [5' -> 3']	Reverse Primer [5' -> 3']
ANX3u9	ANK3	G	A	S	N	cds	GE443	GATGTCATT GAACCAAG	TGTAACTGGC TGAATTGCCTC
AP0D7	APOD	T	C	F	S	cds	GE320	AGAAGGCTCAG AATGACAACC	CAGCATATTA CATTTGAACCT GTAG
AP0D8	APOD	C	G	-	-	noncoding	GE371	CAGCTTCTTG TGTGTCCCG	TCGTGTTGAT TGTGTTGTCT
AP0D9	APOD	C	A	-	-	noncoding	GE371	CAGCTTCTTG TGTGTCCCG	TCGTGTTGAT TGTGTTGTCT
AP0Du1	APOD	A	T	E	V	cds	GE304	CTGGCTTCGA CGNCAATG	ACCCAGTCACT CTGGCTCA
AP0Du2	APOD	C	T	D	D	cds	GE320	AGAAGGCTCAG AATGACAACC	CAGCATATTA CATTTGAACCT GTAG
AP0Du3	APOD	T	G	S	A	cds	GE286	GCTTATTTGCT TTCNATGAGTT GT	TGTCTCAGSAA TTTCTCAAGC
AP0Du4	APOD	C	T	S	L	cds	GE371	CAGCTTCTTG TGTGTTCCCG	TCGTGTTGAT TGTGTTGTCT

FIG. 5E

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
AP0D45	AP0D	C	A	T	K	cds	GE371	CAGCCTTCCTTG TGTGTTGCTG	TGGTGGTGGAT TCGGTTGTGAT	CAGCCTTCCTTGCTGCTGAGATTCTCTTGCCTCTCTCCCTCCCAATAATGCTGCTCTCTCTG gtctcagttatgcacatcggcacctgactgagatctctggccacccagcactatgagaactatgcctctg tctatctcgtactgcatcatcaacttttccagtcggattttgcttggaattcttggaacaggaac tctatctcctcccaagaaacagtcggaacttcaaaaatactctgactcttaataacatctgact cctaactctctcccaagacaggtgaactgtcccccagctctctgaacacaggttctaca caagaaaatga tca/gtgcacagacaggtgaactgtcccccagctctctgaacacaggttctaca gggaggtgcacccactccatgttaacttctgcttgccttccctaccacccccccataaa GACAAACCAATCAACACGA	410
AP0D6	AP0D	T	A	I	H	cds	GE371	CAGCCTTCCTTG TGTGTTGCTG	TGGTGGTGGAT TCGGTTGTGAT	CAGCCTTCCTTGCTGCTGAGATTCTCTTGCCTCTCTCCCTCCCAATAATGCTGCTCTCTCTG gtctcagttatgcacatcggcacctgactgagatctctggccacccagcactatgagaactatgcctctg tctatctcgtactgcatcatcaacttttccagtcggattttgcttggaattcttggaacaggaac tctatctcctcccaagaaacagtcggaacttcaaaaatactctgactcttaataacatctgact cctaactctctcccaagacaggtgaactgtcccccagctctctgaacacaggttctaca caagaaaatgcctcagacagaggtgaactgtcccccagctctctgaacacaggttctaca gggaggtgcacccactccatgttaacttctgcttgccttccctaccacccccccataaa GACAAACCAATCAACACGA	410
ARD10	AR	T	G	G	G	cds	GE655	CGGGGTAAAGG AAGTAGGTG	CTCTGCTCAGG ATGCTTTTA	CGGGGTAAAGGAAAGTAGGTGgaagattcaagcaagctcaagatggagtgcaagttaggctggg aaggtctacctctggccgctcccaagactctccagagagatttccagaaattctgttccagagc tgccgaagtgatccagaacccgggcccagccacagagccgagcagcagcagcagcagcagcagc tgccgaagtgatccagaacccgggcccagccacagcagcagcagcagcagcagcagcagcagcagc ggcaggttgctgctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc gcagc tctcccaagcccatctagagcccccagcaggttaccctggctctctgga tgaagaaacagcactc acagccagctcggccctggagtgccacccagagaggttgcttccagagagctgactcagctgac tgccgcagcagcagaggtgctgcgcagcagctcccccagcttaagcagctgctcgcgtgacttAA ccatcaacgttgctcctgctgggcccacttccccggcttaagcagctgctcgcgtgacttAA AGCATCTCTGACGAGG	602
ARD11	AR	G	C	E	Q	cds	GE655	CGGGGTAAAGG AAGTAGGTG	CTCTGCTCAGG ATGCTTTTA	CGGGGTAAAGGAAAGTAGGTGgaagattcaagcaagctcaagatggagtgcaagttaggctggg aaggtctacctctggccgctcccaagactctccagagagatttccagaaattctgttccagagc tgccgaagtgatccagaacccgggcccagccacagcagcagcagcagcagcagcagcagcagcagc ggcaggttgctgctgctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc gcagc tctcccaagcccatctagagcccccagcaggttaccctggctctctgga tgaagaaacagcactc acagccagctcggccctggagtgccacccagagaggttgcttccagagagctgactcagctgac tgccgcagcagcagaggtgctgcgcagcagctcccccagcttaagcagctgctcgcgtgacttAA ccatcaacgttgctcctgctgggcccacttccccggcttaagcagctgctcgcgtgacttAA AGCATCTCTGACGAGG	602
ARD12	AR	G	A	B	E	cds	GE659	CCACTTTCCC GGCTTA	CGGCCAGGCC AGTGGGA	CCACTTTCCC GGCTTAAGCagctctcgcgtgactttaaagacatctctgagcagcagcagcagc atgcaactctctcagcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc ggga tga/ga/gcctcgggggctcccaactctcccaaggaacttacttgggggagcacttgaccatt tctgaacacgcgcacaagaggttgctgaagcagctgcgggtgctcagctgggctggaggaggtt ggagcactctgagctccggggagcagcttccggggat tgaatgacgcgcctccctctctgagagct caccgctgtgctccactctctgtgctccat tggcgaatgcaaggttctctgctgacagcagc agc agc tgcgctcactctctctctcagagctcagagctcagagctgagcagcagcagcagcagcagcagc gactactacacttTCCACTGCTCTGGCGG	616

FIG. 5F

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')
ARd9	AR	C	T	L	L	cds	GE573	CATGCTTCCC TCGCCA	CAGCGTTCTC CCCTGTATAA
ARu1	AR	G	A	K	K	cds	GE584	TTCAGTGACAT GGTTGCAT	AGTGTCTCTC CTCGAAAGTGA
ARu2	AR	T	A	L	H	cds	GE575	AAACTTCCCC CATTCCTMTT	CAAPAGTGTG CTCTCTGAATC
ARu3	AR	G	C	-	-	noncoding	GE586	TCAGACTTAGC TCNACCCTGC	CCAAGCTGCTG TATTTTAGTGA G
ARu4	AR	T	A	L	.	cds	GE630	TTTTGACCACT GATGTAAATTT C	AAATATGATCC CCCTTATCTCA
ARu5	AR	T	G	S	A	cds	GE1221	TTCATGTGGTA GGNTAATTT CA	TCGTGCTTAAA GAGAGACTAGA AAAT
ARu6	AR	T	G	C	G	cds	GB575	AAACTTCCCCT CATTCCTMTT	CAAAGTGTGTC CTCTCTGAATC TC
ARu7	AR	T	C	S	P	cds	GB659	CCACTTTCCCC GGCTTA	CGGCAGAGCC AGTGGGA

FIG. 5G

SUBSTITUTE SHEET (RULE 26)

Poly Id	Gene	ref NT	alt ref AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')
BDFU1	BUNP	A	G K R	cds	GE1184	AAGCCTTAACC AGTTTTCG	GTTCCTCTCT GGTCATCGA
CD36J8	CD36	G	A P	cds	GE459	TGCTTAAACA GTGACTTTGT TTTTGT	TGCATTCA TTTGTTACT
CD36U1	CD36	A	T Q L	cds	GE437	CTGCTGTTC TTAGAGTTCC	TGTTGTTGTC TACTCACTGCC A
CD36U2	CD36	T	Q V V	cds	GE437	CTGCTGTTC TTAGAGTTCC	TGTTGTTGTC TACTCACTGCC A
CD36U3	CD36	G	T V F	cds	GE459	TGCTTAAACA GTGACTTTGT TTTTGT	TGCATTCA TTTGTTACT
CD36U4	CD36	T	A G	cds	GE437	CTGCTGTTC TTAGAGTTCC	TGTTGTTGTC TACTCACTGCC A
CD36U5	CD36	G	T V V	cds	GE476	TGGAATGCAC TCTTTTT	ATGACCTGTC TACTTGAGG
CD36U6	CD36	G	A S N	cds	GE426	TCCTAGGATC TGCTCATTTG	CTGTGATGAC ACAAAACA
CD36U7	CD36	A	T V V	cds	GE440	ATCATCTATA CTTCAGTATCT CTTCATCT	TGCATTCA TTTGTTACT

FIG. 51.

[illegible]

FIG. 5J

FIG. 5K

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
CYP11A u3	CYP11 A	G	A	E	K	cds	GE585	CTGCAGGGAAC CTCAGCTCTT	CACAGGGGCA ACAAAGGT	CTGCAGGAACCTCACTCTTTTtctctctctccccacagctgacatataccaccgaactt ctactgggaattgagacagaagaagtgcttcaacagattaccggcagatctctctacagactcc tgggagacagcagaagatgctctc/g/a/jaggacataaggcccaacactcagagatgctgcgagg aggggaggacacgg/gaggtggctgtagggcggcACCTCTGTCTCCCTCTGTC	248
CYP11A u4	CYP11 A	T	A	L	Q	cds	GE556	CTACTCCCCAC CAGAGC	CCAGGATTTGG AGTTGGG	CTACTCCCAACAGACGctccatgacctgagtggaactgtatagatggcagcgaacctgaag gtgcaggaatgctgcggggagaggtcttgctgcgcggccagggccagggagacatggccac gatgtacagctggctccctctctcaaaagccagcatcaaggagacat/a/jaaggcaagccac aaccacccatgcsGCCCACTCTCAATCCCTGG	229
CYP11B la30	CYP11 B1	A	T	N	I	cds	GE570	TCCAGGACCA AAGTCTGAG	GGCATCACCTT CTCTGGGT	TCCAGGACCAAGTCTGAGggctgctccgctccccggataggagcaactgtatccagaaa tctatcagaaatggctctcagccgcccctacagtcacacagctatgctggggagctccgttg a/a/l/lkpcgaaactgtgcagatgcca/aaggcacaactataggaaactcactgcggaggcg tggacacgctcagggcggaaccagcgccACCCAGAGGGGTGATGCC	243
CYP11B la31	CYP11 B1	A	G	-	-	noncoding	GE577	GAATGGCCCTG AATGGC	CTCCAGGCTCT CTGAGCGTG	GAATGGCCCTGAATGGCGctctcaaccgattgcggctggaactctccagagctctgaagaag tgcagagttctctccagatgggtgagcagggcgccagctgctccagagctcttccactacacat gctgcagaaacgcccggggagctgacctggagctccagccagcactcttccactacacat tggacacgctggggccac/a/g/lgggaagatcACGCTCTAGAGACCTGGAG	246
CYP11B la32	CYP11 B1	T	-	-	-	noncoding	GE617	ATGGCACTCAG GGCAAA	AGGGCTCTGGG TGTTCCC	ATGGCACTCAGGGCAAGgacagagtgctgcatggcagtgccctgctccctgcaaaagggaca ggcactggggcacagagacgcccgggtctccagagacgtgctgacctttgagcactgcccagc gtccaggcaacaggtgctgagctgagctgcagatctggaggaggaggttatagagacgtgcac ctggaagtacaccagactctccaggaactggggcccatcttcaggtaaagctccctctggccc/a /t/lcgtGGGAACCCAGAGCCCT	285
CYP11B la33	CYP11 B1	C	T	-	-	noncoding	GE625	GGAGGCAGCC GGAGGC	GTGTCCCTTCC CCATAGCAC	GGAGGCAGCCAGAGGC/c/lcgggggtgctctgtgtcagcagtgcatctcccgaagccag caacttgctcttttggagagcggtgggctgtggccacagcccaactgtctgcccagctg aactctcatgctctggagtcagctcaaatccacactccagctcatgtatcatgccaggag ctgtctgctgagcagcagcccaaggtgtggagagcactttgaagctgggagctgcatctcc agtacgggtggccagggagcccggtcagTGCTATGGGGAAGGGACAC	307
CYP11B la34	CYP11 B1	G	C	-	-	noncoding	GE1231	GCACCTTTGAG GGTCTGAGAA	CTTGGGGTGAG GCAGAAA	GCACCTTTGAGGCTCTGAGAGgctgcacacgtcga/gggctgcgacagcagatggaaac cc/g/lcgtgtgacacaggtgctgaaacacactccaggtggagacactaacccagggagaca aagatggtctacagctaatatgagggccacagcatgttcccctcttccacttcagagccatcaa ctaatcacgtctctgcaccaggggtccagcctggccaccagcctcccttTCCTGCTGACCCAG G	261
CYP11B la25	CYP11 B1	T	C	Y	Y	cds	GE570	TCCCAGCACCA AAGTCTGAG	GGCATCACCTT CTCTGGGT	TCCCAGCACCAAGTCTGAGggctgctccgctccccggataggagcaactgtatccagaaa tctat/l/c/laggaaatggctctcagcgccctcaacagtcacacagctcgtggcgaggctct gttgaaatcggaactgtcgccagatggcaatggcaatccagcccaactataggaaactcactgcaggaggcg tggacacggttcaggcgggcaaccagcggccclCCAGAGAGGGGTGATGCC	243
CYP11B la26	CYP11 B1	C	G	N	K	cds	GE577	GAATGGCCCTG AATGGC	CTCCAGGCTCT CTGAGCGTG	GAATGGCCCTGAATGGCGcttcaaccgattgcgggtgaaatccagaaagtgtctgcgccaa lctgctgcagaggttctcccgatggatggatgcagggaggttctccagggcctctgaaagaa gaaggtgtgcagaaagcccccgggagagctgacctggagcgtccagccagcatcttccactca tcatagaaaggtgtggggccacatgggaagatccACGCTCTAGAGACCTGGAG	246

FIG. 5L

[illegible]

SUBSTITUTE SHEET (RULE 26)

[illegible]

SUBSTITUTE SHEET (RULE 26)

[illegible]

[illegible]

FIG. 55

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
DRD1u5	DRD1	A	C	A	A	cds	GE1173	GTGACCCCTAT TCCCTGCTT	CTCTCCAGGC GCGCAATG	CTGACCCCTATTCTGCTTAggaacttgagggtgtcagagccctgatgtgcttctctttagg aagatgaggactctgaacacctctgcacatggagggagctgggctgggtgggtggagaggaactctc tgtctgatctcctcactgctcttctctgtcgtcactcctctccacgtctctcctggggaacacgc tggtctgtgctgcagttatcaggttcccgacctgcggtccaggtgacaaactctcttctgcaac tacttggtgtgctgactctgtgtgtccgtctctgtgtcatcccttggaaggcagtggtgtgagat tgctggctcttgcccttgggtctcttgtaacctgtgtgggtggcccttgacacatgtgtgctcca ctgca/cjctccatctcaacctctgtatcagcgtgggacaggtattgggtctatctccagccc ttctcggatagagaaagatgaccccaaggagcctctcatctgcatcgtatgggtggacagacct tgtctgtactcatctctcaccagtgagctcagctggccacaaaggcaaaccccaacagccc tctgatgaaatgccactctcctgtgagaccatagacaactgtactcagctcagctcagcaggac atatgcatctcatctctctgtaataagcttttaccacctgtggccatcatgattgtaccctaca ccagatctcacggattgctctgaaacaaatcacggcgCATTCGGCCTTGGAGAG	770
DRD1u6	DRD1	A	G	K	E	cds	GE1170	GGATCTACAGG ATTGCTCAGAA AC	GCAATCTCTC TAGCTTTTGG	CGATCTACAGGTTGCTCAGAAACaaatcacggcgcatctggcctctgagagggcgccagctccac ggcaagaattgcccagagaccaccacaggttaatggaaagctctcgaaatgtctctcaacgggaagtctc tttaagaattctctcaaaagagaactcaaaagctcgaagacctctcggtgtgtatctatgggttgt ttgtgtgctgtggctaccttctctcacttgaaactgcatcttggcctcttggtgggtctggaatg acgcacctctctgcatctgattctcaaacaccttgacctgttttgggtttgggtgtggaatgaaatc atcttgaaacccatctatctgaccttaatgtgattctcggaaggcatttctcaacacctcttag gatgtcacagacttggcctgcgagataatgccatagagcgtgagctgaatcaatacaataggg ggcggatgtttccagccatcatgacggaggtcccatctccaggagtgccacatctcgtgttba cctgatccacatgctgtggctctctgaggacctgaaaaaggagggagggcagctggcactgcaca gaccttggag/a/gjagctgtcccaagcctatctggtcatattggactatgacatgagctgacgtctc tgggagaagatcccaacctacacaaacgggtcagaccacacacacacacacacacacacacacac tgccacacatgctctcatcccaaaagctTAGACAGAGATTCG	753
DRD1u7	DRD1	A	C	R	R	cds	GE1173	CTGACCCCTAT TCCCTGCTT	CTCTCCAGGC GCGCAATG	CTGACCCCTATTCTGCTTAggaacttgagggtgtcagagccctgatgtgcttctctttagg aagatgaggactctgaacacctctgccatggacgggagctgggctgggtgggtggaggggactctc tgttgatctctcactgctcttcttctgtcgtctcatctctgtccagctcagccttggaaggacacgc tggtctgtctgctggttatcaggttccgja/cjcaacttggctcacaagctgaacaaactcttttgt catctcttggtgtgtcagatctcttggtggcctctggtcactggcctggcctggagcagtggtgtg agattgctggctctggccctcttggtctctctgtgaacatctgggtcagctttgacatcatgtgc tccactgcatcactctcaacctctgtgtgacagctgacaggtattgggtatctccagccc ttctcggatagagaaagatgaccccaaggcagcttcatctgtcatgtgtggctgagcagacct tgtctgactcatctctcatccagtcagcgtcagctggccaaaggcacaacacacacacacacac tctgatgaaatgccactctcctggctgagaccatagacaacttgacacacacacacacacacacac atagcactctcatctctctgtaataagcttttaccactctgtggccatcatgattgtcactctaca ccagatctcacaggattgctcagaacaaatcacggcgCATTCGGCCTTGGAGAG	770

FIG. 5Z

[illegible]

FIG. 5AA

[illegible]

FIG. 5CC

[illegible]

FIG. 511

[illegible]

FIG. 5QQ

[illegible]

FIG. 5UU

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' - :')	Reverse Primer (5' → 3')
F5d47	F5	G	A	-	-	noncoding	GE949	TCCCTAACAT GGAGTTTACT TT	GCACACTGTA GCCGGTA
F5d48	F5	A	G	-	-	noncoding	GE316	GCAAGGTATT AATCATCTTC T	GCACAGTCTC AGATTGCTTT
F5d49	F5	C	G	H	Q	cds	GE387	GTGCCCCAGA GGAACAATA	TCATATGGCTG AGTTCTGGAG
F5d50	F5	A	G	T	T	cds	GE389	GAAAGTAGCT ATGAATAATC CAAGA	GGGCTTGAAAT GGCGAATGT
F5d51	F5	A	C	E	A	cds	GE496	CTTTCCTCAGA CCTTGTGTGAG	AAGGGAAMC TGGCCNMC

FIG. 5WW

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
F5d52	F5	C	A	-	-	noncoding	GE910	GCAGGAGATATGGTG TGGTGTG	TTTACTGTATAC AATGCGAGGT T	GCAGGAGATATGGTGTAATAATCA(c/a)ggaaccaccacatgacggatgtccacacc ccctggatcggaaatggaagatgaaacacacacagctctctcgtttaagaatcttc gggtggagattactcgggaacctccctccgcccgtctgaatgcccgaggtgatgctgg caagcaaggtcaagtatactctatggtgtctctctcgggctcccaaaaaaagcagcc ctcactaccacaacatggaggatctggagcaggatatttaactctctctgcaaatcaaa atcctaggctccaanagtcaggatgctggagcttgatantttacttataacttctctaaa aanaagtcattatctctctgtgtgttttggcttattctgttttaaaccaataaagaacaa taatttttgcgaaggatanaatgtaaatattttagctcttggggccatcacattctgcca taactactcAACCTGCCATTGTATCACTAAA	552
F5d53	F5	C	T	-	-	noncoding	GE915	TGATTATCAGA AGAGCAAGGAA A	ACTGTGACCA GTGTGATTTA	TGATTATCAGAAGAGAAAtctcggagaaggaggaatacaattactctgtttttccagg aanaagataatctactcaggcttgataggctccctctcaatctgcgaagggaata-tacataag gacagcaactatcctatggacatggagaaatttcttactatttatgacctttgatgaaagaa ggctgggtactatgaiaaagagtcacgaagttcttggagactcacatctcagaaatgaaaaat cccatgggtttcacggtatttctct(c/a)ggactttgattctaatctctaatTAATTCACACTG GGTCACAGT	334
F5d54	F5	A	G	-	-	noncoding	GE915	TGATTATCAGA AGAGCAAGGAA A	ACTGTGACCA GTGTGATTTA	TGATTATCAGAAGAGAAAtctcggagaaggaggaatacaattactctgtttttccagg aanaagataatctactcaggcttgataggctccctctcaatctgcgaagggaata-tacataag gacagcaactatcctatggacatggagaaatttcttactatttatgacctttgatgaaagaa ggctgggtactatgaiaaagagtcacgaagttcttggagactcacatctcagaaatgaaaaat cccatgggtttcacggtatttctct(c/a)ggactttgattctaatctctaatTAATTCACACTG GGTCACAGT	334
F5d55	F5	A	G	G	G	cds	GE923	GATCATTCCTT TTCTAGGTT	TTTCAGATTACG AGGTTAGGGGA	GATCATTCCTTTCCTTAGGTTTcgttttaaaaatttagcatccagccgtattctctacatgccca tgg(a/g)gctttctctatgaaanaatcatcagggaggaaga(c/g)ttatgaagatgactctctgaatgg tttaaggagataatgctgttcagcaaatagcagttatacctacgtatggcatggtccactgagcg atcaggccagaaagccctggctctgctgctggcttggctactactcagctgtgaacccag taggactttcatgaaagtttttctcatTCCTTAACCTCGTAACTGAA	311
F5d56	F5	C	G	T	S	cds	GE923	GATCATTCCTT TTCTAGGTT	TTTCAGATTACG AGGTTAGGGGA	GATCATTCCTTTCCTTAGGTTTcgttttaaaaatttagcatccagccgtattctctacatgccca tggactttctatgaanaatcatcagggaggaaga(c/g)ttatgaagatgactctctgaatgg tttaaggagataatgctgttcagcaaatagcagttatacctacgtatggcatggtccactgagcg atcaggccagaaagccctggctctgctgctggcttggctactactcagctgtgaacccag taggactttcatgaaagtttttctcatTCCTTAACCTCGTAACTGAA	311
F5d57	F5	G	A	-	-	noncoding	GE924	TTGCTTGCCT AAACTCTTG	GAANACAGAC CGAAMATTAC TA	TTGCTTGCCTTAACCTTTCGttatccaggaaggaatgagn(g/a)lnttttatctgcagtgcc tactgaanaagctcttctctctcataggagcctaactgaggtgggagacacagagcgttga caagcaatcgtgctactatttctgtgtgttgaagcaagcagcgtggagccagctcatctccc taatgtacacagtcantggaatgtaattggaatcgaatggaatggaatggaatggaatggaat taaacacagtaaaatcatTAGTAATTTTCGTTCTGTTTC	303
F5u1	F5	C	T	N	N	cds	GE173	AAATTTGATT AACTTTGTAGA TCGTG	GGATTCAAGTAG AAGTAAAGAT TCAA	AAATTTGATTACTTTGTAGTACTGTCtcaaaaattggccagccgccccatagcatttaacc tcaggaagtcacctctctccttatgaagatgaatca(c/t)ctctcttccactcaggtttg AATCTTCACTTCTACTGAATCC	153

FIG. 5XX

[illegible]

FIG. 5YY

FIG. 5AAA

59/178

[illegible]

FIG. 5BBB

[illegible]

FIG. 5DDD

[illegible]

FIG. 5EE

[illegible]

FIG. 5FFF

[illegible]

FIG. 5LLL

SUBSTITUTE SHEET (RULE 26)

[illegible]

FIG. 5000

74/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
FGb9	FGb	C	T	Y	Y	cds	GE392	AAGGGAAGAA GGCAGTTTT	CCACGGAAGTG GTAGCTATTAA A	AAGGGAAGAAAGCAGCTTTTtagtttccccaaattttatttttggtagagattttattttttttt tctcttttagtgtaatttgctctgggaattgataaaattagccagcttaccagatgggagccaca gaactttttgatagaattggagactgaagaggagacaaagagcttccactatggagatttcac tgtacagaatgaagccaaacaaatccagactctcagtgaaacaaata[c/t]agagggaacagccggt aatgacctcatggatggagacatctcagctgtaggggaaacacagagaccatgaccattcacaacgg catgtcttcagcacgtagacagagacaatgacggctggtagtggtggcacttttgcctctgc tttaaaatcacactaatatcatctactcagaatcatcaaaatattttTAATAGCTACCACTTCC TGGG
FGd3	FGG	A	G	M	V	cds	GE337	TTGGTAATAGA CAGCTCTTCAT AGACT	GCAGTTAAATT TCTACAAATCA TCC	TTGGTAATAGACAGCTTTCATAGACTTgcagagggtaaaagattccagaaataatgtagtaca ttctacgacttggtaggtggcactctcacaagcatctactctctactatggttatgataatggca ttatttggccacttggaaaccccggtggtatctccatgaagaaaaccact[a/g]tgaagataaat ccattcaacagactcaatatttggagagacacacacccctggggggggccaaacaggtcca gaccagagcacctctggggaacagaaatgactcaactttaccctgaGANTGATTGTAGAAAATT AACTGC
FGd4	FGG	A	G	-	-	noncoding	GE349	AAATACCTTAG CAGTTTCCAAA GMAA	TGGGTAGCCAC TTTCTTAACTA TTC	AAATACTAGCAGTTTTCAAAAGAAATataaaattactcttctgaaagggaatacttttttgt ctcttatttttctatcttatgtttctgtttgttagatatttgcaggaataataataaattcaata atcaaaagattgttaaccggaagaggagggtagcccgcttgaacacagctgcaggaaccttgc aaagacaggtgcgaatccatgatatactctgggaagggtaactctgtagaggttatattgggatta ggttcatcaagtaagtaattgtaaaaggagaagtatgactgg[a/g]aagtatagGAATAGATT AGAAAGTGGCTACCCA
FGd5	FGG	G	A	G	R	cds	GE360	TGCTGATGTGA AAAGTAAGAAA AT	CAAGGTGCTTA GAAAGTATCT GC	TGCTGATGTGAAGAAATAGAAATattcttggaaaatgaatagtttactatcatgllanaagcta tttttcaaggctggcagactcttactctcttcaaccacagtaaaagttgactctctctctctct agattgtcaagacatttgcataaagaggatttggacatctctctactctggaacacagaaattctg ctgaagctcaaccagcaattcttactctactgtaaatgataatgggtctgtgaaatggatggactgt gtttcagaaggttaatttttccccccatggtatttaaataattctctactatgtttctgacctata tgGCAGATACTTTTCTAGCACCTTG
FGd6	FGG	T	A	-	-	noncoding	GE372	GAACCAAGCT CTGTATTTTGG AC	CCATTGTCTAT TGATACTTGA AAG	GAACCAAGTCTCTGTATTTTTGACaaaatgptgacagcatctctttaca[c/a]gcaattgtag tctatttctctcttctgtcttgcataatgttaatttagagacttgatggcagtgtagatttcaag aaaaactggattcaataataaagaggatttggacatctctctactctggaacacagaaattctg gctgggaatggaagatttcaattgttaagacacagctctgcatctccatctcattgcatgaaggtgg aactggagactggaatggcagaaacaggtactgttttgaatgacttccaaactttttattgttaa agattgctgggaatgtagcttttCCAACTTCAATAGAACATGG
FGd1	FGG	A	T	Y	F	cds	GE404	CATCTACGAA ACAGGGAATT T	TCCACTTCCAG TTTCAAGAAC T	CATCTACGAAAGAGGAACTTctggagatccctggaggggtcagcatgtgattgttatttcc ttctctcagtagctcagact[a/t]tggcaatgttcaaggtgggagctgaactgaactgaactgacc cctaactatgctactctcggtgggtggtgtagatgctcttggatggttgaatggttgaatttggcg atgactcagtagacaagtttttcaatcccaataatggcatgactcagtagtctgtagtgaacagtgca aatgaatgttgaagggaactgtgctgaacaggtgatctgtgttggtagtgaacagtgctga cgctggccatctcaatggatttatacccaaggtatgttttctcttcttagatcccaagttcaatg tatagtgtaactattttcataaaaaataatagataagagaataatgaagaataatattataa agatagtagggattttatcatgttcttcttcttcaactaagtTCTTCAACTGGAAGTGA

FIG. 5QQQ

[illegible]

[illegible]

FIG. 5EEEF

[illegible]

FIG. 5GGG

FIG. 5HHHH

FIG. 51111

FIG. 5JJJJ

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')
HSD3B2 u15	HSD3B 2	C	T	-	-	noncoding	GB665	CTGCTGGAAA TAGTGACCTTC	GCCTCTTTTGT TGAACGTGTGTG AA
HSD3B2 u16	HSD3B 2	A	C	-	-	noncoding	GB665	CTGCTGGAAA TAGTGACCTTC	GCCTCTTTTGT TGAACGTGTGTG AA
HSD3B2 u17	HSD3B 2	G	T	L	L	cds	GE1194	CAGAAGAATGC ACCCTGATC	GCCAGATCTCG CTGAGCC
HSD3B2 u18	HSD3B 2	C	T	L	L	cds	GB665	CCTGCTGGAAA TAGTGACCTTC	GCCTCTTTTGT TGAACGTGTGTG AA

FIG. 5KKKK

FIG. 5LLLL

[illegible]

FIG. 5MMMM

[illegible]

FIG. 5NNNN

[illegible]

FIG. 5PPPP

[illegible]

FIG. 5TTTT

[illegible]

FIG. 5VVV

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
HTRIEL u4	HTRIE L	A	T	E	V	cds	GE1160	TTCCCTTTGTTA CAGGTATCCAT	GTGGGATGAG AAGCTCCA	TTCCCTTTGTTACAGGTATCCATTTtcagctatataaacttttaaaacaaagaagaatggatttc ttaaattcatctgatcaaaacttgacctcaggaagactgttaaacagatgccatccaaatctc gggtccctcaactctgtctgggtcgtgactgatgaacaaactatacctcccttgatcgctg caattattgtgacccggaagctcaactatccagccaatttttaattgttcccttgagtcaca gaattctctggcgtctcgtggtgacccctcaagcatttgtatattgtgagagagcgtggt tatgggcaagtggtctgacatttggctgaggttgatcactctgctgcagcgtcccaact tgcatctctcagctatagctttggatcgtgatacagagatcagagatgtgtt[17]lgaagc caggaagaagactcacaagcattgctggcattatgatacaagtttggattatactctgtttta tctctatgctctctctatctggaggaccagaagactagcagagatgatgaatcatcatcaag cacgaccacattgttcccccatttctcaaatTTGGAGCTTCTACATCCAC
HTRIEL u5	HTRIE L	T	A	I	I	cds	GE1158	CGACCACTTG TTTCACCA	AGTTATTCTC CCCTCAAA	CGACCACTTGTTCACCACTTactcaacttggagctttctacatccccctggcattgattt tgaattctactcaaaaataataagcagcagcaagcattatccagaagcagcagcaagatgg atttgcgaaggagagtgaaatggccctgctcttttggagagtggtggagaagaagcactaaatcgt ttccacatctctatgactagaagtttttatctgaccctatcaacagattttgacaaat[17]a catgcacagtgagaagctcaggtctgaattcaagcatgagaactcttggagagagcgaagaat ctcaggtacaagaacaggaagcagcactacctgggatttaacttgggtgcatttggtaatat gttggctctcttttttgaagaagaattagtttgaatctgtgacaaatgaaatttctgaa gaaatgtcaatttttggcatggctgggtatctcaattcccttataaactcactggatttcaac aatctttaaagagactcagaagaagcattccaaagctgtgcgagtgcgattgattttaa aatgttattattagaagatgggggtttttcagggaggaatNACT
HTRIEL u6	HTRIE L	T	G	D	E	cds	GE1160	TTCCCTTTGTTA CAGGTATCCAT	GTGGGATGAG AAGCTCCA	TTCCCTTTGTTACAGGTATCCATTTtcagctatataaacttttaaaacaaagaagaatggatttc ttaaattcatctgatcaaaacttgacctcaggaagactgttanaagaa[17]gacatccaaatctc gggtccctcaactctgtctgggtcgtgactgatgaacaaactatacctcccttgatcgctg caattattgtgacccggaagctcaactatccagccaatttttaattgttcccttgagtcaca gaattctctggcgtctcgtggtgacccctcaagcatttgtatattgtgagagagcgtggt tatgggcaagtggtctgacatttggctgaggttgatcactctgctgcagcgtcccaact tgcaattcagctatagctttggatcgtgatacagagatcagagatgctgtgagatgcagg aaaaggaactcagaagcattgctgcatatgattacaatagtttgattatctgttttctc tatgctctctctatctggaggaccagaagactagcagagatga[17]lgaatgcatcatcaag cacgaccacattgttcccccatttctcaaatTTGGAGCTTCTACATCCAC
HTRIEL 1	HTRIE	T	C	I	T	cds	GE1157	CAGCCAAAGGA AATAACCA	GCACCCAGCTT GGAGTAAAT	CAGCCAAAGGAATAACCAAGcttccacagtgagctgagaacagggagaacatgaacat cacaactgtacacagagagccagatgggtataagaacaaagacactactgagaagatgctca tttgcatactctgtgtgtcatcaccacccctcaccagctgtgtgaacttggctgtga[17]clcat ggctattggcaccaccaagactcaccagctcccaactactcaactctctctctcgtcgtg cggacctctgtgtgagcattgctgtgaggtgtgagtgatggacatgactgtcgaactgtccat aagcttgggtactctctgtgggtgtgggtgtgagtgatggacatgactgtcgaactgtccat cctcactctgtcatctggcctggaagagactgggcaatcccaaatgctattgaaacgcca ggaaaggagcggcgaaggcggtgatacttacccttggaccactctcatttcaatc tccatgccccctctgtctggagaagccacccgctccctgaagcctccctctagtcgaccat ccagcagcaccatgttatcaactTTTCTCCACCTGGGTGC

[illegible]

FIG. 5YYYY

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
HTR2ad 14	HTR2A	C	T	-	-	noncoding	GE1167	TGGTACTGCGA AACCAG	CGCAGCTGCTAG GATCTCTGTT	TGGTACTGCGAACCAGTCTTCTCTACCACTGAGGTTTGTGAATAATATCTGGTGGGCTA TTCTGTGGAAGAAAGAAAGCCGCTTCAAGTGTGATATATTAATAAATAAGTGTCTAATAGTT TATCAGAGTTATACCCACAGCTGCTAGCCCTGAGCTGCTATGTGGCAATGTGCTAGTAAATCTCC ACTCTGGACACAAACACTGTGGCTTGGATGGAGTGCAGACACTCACAGCACTC/C/T/GAG GACATGTTTCTTAGCCTTATTTATGTTCTCTCTCAACCTCAGCTCAAAAATATACCAAG TGCCACTACCTACCTTAAGGGAATCTGACAAAGAGCCATCTCCAGTTCTCTCAAGAGCA AGGAAACTCTCCCAATATTAATATGACAAAGAGGAGAGAAAGCCCTGTTGGTCCGAC CCTCTGGCTGTGCTACCTCTGAGCTATGACATACACATAGAGGGAGGCTGATGAAAGAAC GAGACAGTCAGAGAGTACTCCTCTCCGTTGAAACAGGAGTCCCTGGTGGACAGACTCTTC CTACTTCCCATGCACTTCTTGTGGACTTGGAGGCTGCTGATGATTTCTAATGTTGTC CTGTGGAGGAGCGCACAGGGGAGGAGAACCCAGCCGAGCGTGCAGAGGAGCCAACAGG ATCTAGACAGTGG
HTR2ad 15	HTR2A	G	A	-	-	noncoding	GE1169	CGGTGCGAGAG GAAGCC	CTCACCAAGC GAGGACAA	CGGTGCGAGAGAGCCAGTCTCTAGCAGTGCAGCTGCTCAGCTCTTGCATCAGTTT TTGAAGTCAGAAAGAAAGAACTCAATCTCATATATATGCTGTGAGAGTCAAGAGAGGG GACTCTACCCAGTTTAAATCTGAGAGATGCAGCGAGTCAAGATACCAATGATCTCAT GTGTGAACCTGAGACAAATGTAACTCTCATGCTGCTATATTTATGCTGTGAATTTCTT TCGGTTTGAATCATCATCTTGGCCAACTGATCAATCTTCAATGAGATCTCCAGGAGGAAAGT TGCTGTAATTTAATTAAGTCTTTTGTCTCTTCTTATAGCTAAGCAATATATAGGAG CTGAATCTGACAGAGCTGAGGCAATCAGCTGAGAGTGGCTGAGAGTCTGTAACCAAGAG TACATCAATTAATAGGATACACTGATGATTTTAAATGACTTCTTAACTAGATGAT GTACATCCCATGTTCTGATTTGCTGCTATTTAAATGACTTGTCTAAGTACCATCTC G/AJGCAATACCAACAAATGATATAGTTAAACAAGAGTCCCAAGTGTATATAAATCTTCTT CTTGTCCAGACATTTATCTTCCGGAAGCTCAAAAAAAGCCCTGCAACCTCTATGCTAAAG TTTCTATCTGCTTTTGTCTGCTGTTTGTGAG
HTR2Au 1	HTR2A	G	C	-	-	noncoding	GE1167	TGGTACTGCGA AACCAG	CGCAGCTGCTAG GATCTCTGTT	TGGTACTGCGAACCAGTCTTCTCTACCACTGAGGTTTGTGAATAATATCTGGTGGGCTA TTCTGTGGAAGAAAGAAAGCCGCTTCAAGTGTGATCTATTAATAAATAAGTGTCTAATAGTT TATCAGAGTTATACCCACAGT/C/ACTGCTAGCCCTGAGCTGCTATGTGGCAATGTCTAGTAA TTCCACTGTGACACAAACACTGTTGGCTTGGATGGAGTGCAGACACTCACAGCACTCCGAG TGACATGTTTCTAGCCTTATTTATGTTCTCTCTCTCAACCTCAGCTCAAAAATATACCAAG TGCCACTACCTACCTTAAGGGAATCTGACAAAGAGCCATCTCCAGTTCTCTCAAGAGCA AGGAAACTCTCCCAATATTAATATGACAAAGAGGAGAGAAAGCCCTGTTGGTCCGAC CCTCTGGCTGTGCTACCTCTGAGCTATGACATACACATAGAGGGAGGCTGATGAAAGAAC GAGACAGTCAGAGAGTACTCCTCTCCGTTGAAACAGGAGTCCCTGGTGGACAGACTCTTC CTACTTCCCATGCACTTCTTGTGGACTTGGAGGCTGCTGATGATTTCTAATGTTGTC CTGTGGAGGAGCGCACAGGGGAGGAGAACCCAGCCGAGCGTGCAGAGGAGCCAACAGG ATCTAGACAGTGG

FIG. 5ZZZZ

Poly Id	Gene	ref NT	alt NT	alt ref AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
IITR2AU13	IITR2A	A	G	-	noncoding	GE1169	CCGTGCCAGAG GAAGCC	CTACCAAMCC GAGGACAA	CCGTGCCAGAGGAGCCAAacagagctctagcagtcggagctggctcagctcttgcgcatgcagttt ttgaagtcagcaaaacagaaacaaatctactatctatctatctgctggtggagagatcaagaagagggg gactccaccagtttcaattatgctggagagatgcagcagagtcacagatacaaatgatactcatc gtggaacccctgagcaaaatgaagttgcatacgcgctatatttctatgctgtaattttctt tcgggttgaatactgcttggccacatgtaatactcaatgaagataccagatggaaggaag tgctgctaatcttacttaagactttttgtttctcttatttagctaaagcattcaacatagagg ctgaattctcagcagcagctggccaatcagcctaagaatggctcaggaactgaaccacaaga tacaaccaattactaggattaaactgtagtattttaaactgactttcttaagttagaatgt gtacatcccactgtttctgattgtagctattttaaactactgttgcataaactagatccactgt gctcaacacaaatgagatagtttaacaaagatcccagtagttataaacttttctctt gtcccaacatttatcttcccagla/gtgcctcaaaaaaaccttcgcaacctctatgctcaaaag tttactctgtcttttTGCTCGGTTTGTTGAG
IITR2AU2	IITR2A	A	G	-	noncoding	GE1167	TGTTACTGCCA AACCA	CGCATCTCTAG GATCTGTT	TGTTACTGCCAAMCCAActtattctccacacatgtgaggttttgaaataatctctgggtggccta ttctgtctgaagaaatgaagccagtttcaatgggtgtctcttataataaaaaagtgataatggt tatcagagttatcacacagactgctgacccactggagctatggtcccaatctcagatatactcc actctgacacaaacacagcttggcttggatgaagtgcia/gtgcacatcagacacagccag gacatactgtttctagcttatttattgtctctctacccttcagctcaaaaaataccacag tgccacttacctacttaatgggaatctgcacaaaaggccattctccagtttctctcaaaagca adggaataacttcccaataataatgtagcaaaaaggaggaagaaagcgtgtttgtgtccgc cctctggctgtgtctaccttgcagcttatgacatacactagagggaggtctgataaataagac gagacagtcagagctactctatccctgtggaacccagggtcccttggctcagcagactcttc ctacttcccatgactttttgtgagactttgaggggtctgtgaatgtaattcaaatgtgtgc ctgtcgagcgcagccgcacaggaaggagaccagccgagcgtgctcagcagggaagccAACAGG ATCCTAGCAGTGC3
IITR2AU3	IITR2A	A	G	-	noncoding	GE1169	CCGTGCCAGAG GAAGCC	CTACCAAMCC GAGGACAA	CCGTGCCAGAGAGCCAAacggatcttagcagtcggagctggctcagctcttgcgcatgcagttt ttgaagtcagcaaaacagaaacaaatctactatctatctatctggttgggaagatcagaagggg gactctacacagttaactctgtgagagatgcagcagatcacagatacaaaatgatactctcat gtgtgla/gtaccctgaagacaaatgaagtctcatgcgctatatttattgtctgttaatt tccttcgggttgaaatactgttggccaaatgtaatacttcaatgagaattccagaggga aegtgtctctaattttacttaagactttttgtttcttttatttagctaaagcacatata ggagctgaattctcagcagcagctggcaattcagctaaagaatggccgagagatcaacca aagatacatccaatctactatgggataaacactgtagtattttaaactgactttttaaagttaga atgtgacatcccctgttctgattgctgactatttaataactgtttgtaaaactgtacc atcggcatacaacaaataatgataatgtaaacaaagatcccagtagtatacaaaactttctt cttgcacagaacttatcttctcccgagctcaaaaaaaccttcgcaacctctatgctcaaaag ttctactctgtcttttTGCTCGGTTTGTTGAG

FIG. 5BBBBB

Poly Id	Gene	ref N ^o	alt N ^o	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
HTR2Au 7	HTR2A	A	G	-	-	noncoding	GE1167	TGGTACTCGGA AACCA	CGCACTGCTAG GATCCTGTT	TGGTACTCGGAACCAACTATTCTCTCCACCATGTGGGTTTGTGATAATATCTGGGTGGCATA TTCTGCTGAGAAATAGCCGCTTCAATGGTATCTATTAAATAAATAAGTCTAATAGTT TATCAGAGTTATCACCACAGACTGCTCCCTGAGCCTATGTGGCAATGTCAGTAATTC ACTCTGGACAAACACTGTGGCTTGTGAGAGTGCACACTCACAGCACTCCGAGGACA TACTGTTTCTAGCCTTATTAATGCTCTTCACTCTCAGCTCAAAATACCAAGTGCC ACTTACTACCTTAATGGATGTGCACAAAGGCACTTCTCCAGTTCTCTCAAAAGCAAGG AAAACCTCCCA/GJATATAATATGACCAAAAGAGGGGAGAGAAAGCCCTGTTTGTCCGC CCCTGGCTGTGCTACCTTACCTATGACATACATAGAGGGGAGTGTGATGAATGAATC GAGACAGTCAGAGAGTACTCCTCCCGTGAACCCAGGAGTCCCTTGGTGCAGCAGCTCTTC CACTTCCCATGCGAGTCTTCTTGTGACTTGGAGGCTCGTGAATGATTCTCAATGTGTGC CTGCTGAGGCGAGCCGCGAGGAGGAGGAGCCAGCGAGCGTGCAGAGGAGCCAAACAGG ATCTAGCAGTGGC	729
HTR2Au 8	HTR2A	A	G	-	-	noncoding	GE1167	TGGTACTCGGA AACCA	CGCACTGCTAG GATCCTGTT	TGGTACTCGGAACCAACTATTCTCTCCACCATGTGGGTTTGTGATAATATCTGGGTGGCATA TTCTGCTGAGAAATAGCCGCTTCAATGGTATCTATTAAATAAATAAGTCTAATAGTT TATCAGAGTTATCACCACAGACTGCTCCCTGAGCCTATGTGGCAATGTCAGTAATTC ACTCTGGACAAACACTGTGGCTTGTGAGAGTGCACACTCACAGCACTCCGAGGACA TACTGTTTCTAGCCTTATTAATGCTCTTCACTCTCAGCTCAAAATACCAAGTGCC ACTTACTACCTTAATGGATGTGCACAAAGGCACTTCTCCAGTTCTCTCAAAAGCAAGG AAAACCTCCCAATATAATGACCAAAAGGAGGGGAGAGAAAGCCCTGTTTGTCCGC CCCTGGCTGTGCTACCTTACCTATGACATACATAGAGGGGAGTGTGATGAATGAATC GAGACAGTCAGAGAGTACTCCTCCCGTGAACCCAGGAGTCCCTTGGTGCAGCAGCTCTTC CACTTCCCATGCGAGTCTTCTTGTGACTTGGAGGCTCGTGAATGATTCTCAATGTGTGC CTGCTGAGGCGAGCCGCGAGGAGGAGGAGCCAGCGAGCGTGCAGAGGAGCCAAACAGG ATCTAGCAGTGGC	729
HTR2Au 9	HTR2A	G	A	-	-	noncoding	GE1169	CCGTGCGCAGAG GAAGCC	CTCACCAACC GAGGACAAA	CCGTGCGCAGAGGAAGCCAAACAGATCCCTAGCAGTGGGAGCTGGCTCAGCTCTGCTGCGAGTT TTGAGGTCAGCAAAACAGAAACAACTACTATCATATATGCTGATGGAAGATCCAGAGAGGG GACTCTACACCGAGTTTAAATGCTGTGAGAGATGCGAGAGTCAAGATAACAATGATCTCAT GTGTAACCTGAGACAAATGTGAAGTCTCATGCGCTATATTTATGCTGTGATTTCTT TCCGATTGGAATCATGCTGGGCAACATGTAATCTCAATGAGAAATCCAGGAGGAGAGT TGCTGCTAATCTTACTAAGTCTTTTCTCTCTTATAGCTAAGCAACATATAGAGG CTGAATCTCTGACAGCGCTGGCAATCA/GJCTGAAGATGGCTGAGACTGTAACCCA AAGATACATCAATCAATGATTAACACTGGATATTTAAATGACTTCTTCAATGAGAG ATGTGACATCCCCCTGTTCTGATGCTGATTTAAATACTGTTGCTAAACTAGTACC ATCGGATAACCAACAAATGATATAGTAAACAGAGTCCAGTAGTTATAAACTTTCTT CTTGTCCAGAACATTTATCTCCGAGCGCTCAAAAACCCCTGCAACCTCTATGCTAAAG TTCTATCTGCTTTTGTCTCTGGTTGGTGGAG	749
HTR2Cu 11	HTR2C	G	C	C	S	cds	GE1126	TTTTTCAGTGT GCACCTAATG	ACTTACCATAA AGAGTACC	TTTTTCAGTGTGACCTTAATTTGCTCTATGGTTGGCAAT/GJTGATATTTCTGTGAGCCAG TACAGCTATGATTAACAGCACTTTCAATCCTCCCTGAGTGGAGCTCTCAATCCCCAGAGGG GTCAAACTGGCCAGCACTTTCAACGCTCATATAATAATCAAGCAATAGGTGGCAACATCCT TGATCATGTCAGTAAGCATGGAAGAGAACTGCAATGCCCAATCTCTCTTCTAATGCTCC TAGCATTGTGATGCTAGTGGGACTACTGTCAATGCTCTCTCTCTCTGGCAATCTTTAT GCTAAGT	332

FIG. 5DDDDD

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
HTR5A 3	HTR5A	C	T	I	I	cds	GE1319	TCTCCAGTAC CCGAGG	GGATTGTGTGAG ATACCA	TTCTCAAGTACCCAGCGGCTCTCTGACCGAGATGGATTACAGTGAACCTAAGCTCTCTT TCTCTCTCAACCCCTCTCTGAGAGCAACAGCAGCTCTGAGCAACCTCTCTGAGGCTTCTGGTGGCGG AGTCTGCTGGAACCTCTCTGAGTCTGGAGTGTCTATCTCACTTGTCTGAGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CCCAACCTGAGGCTCTCTGAGTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG GCTGAGGCTGCTTGTCTGAGGCTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTGCTCACTCAACCTCTGAGGCTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TGGCTCACTCAACCTCTGAGGCTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG AGGAGCTACTCTGAGGCTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CCCACTGAGGCTCTCTCACTGAGGCTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG GCTGAGGCTCTCTCACTGAGGCTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG GGTGGGTATCTCACTGAGGCTCTGGAGTGTCTATCTCACTTGTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG	800
HTR6d5	HTR6	G	A	P	P	cds	GE1148	CCAGGCGGTGT GCGACT	AGGGTCTGGGT TCTGCTCA	CCAGGCGGTGTGCGACTGCGACTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG ACAGCACTATGAACCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TCTCTGCTGCTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTCTCACTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG ACTGAGTCTGAGGCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTGCTCTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTTCACTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TGACCTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG	511
HTR6u1	HTR6	C	A	Q	K	cds	GE1148	CCAGGCGGTGT GCGACT	AGGGTCTGGGT TCTGCTCA	CCAGGCGGTGTGCGACTGCGACTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG ACAGCACTATGAACCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TCTCTGCTGCTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTCTCACTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CCGAGCTGAGTCTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG GCTGAGTCTCTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG ATTCTCACTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TGACCTGAGGCTCTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG	511
HTR6u2	HTR6	G	T	A	A	cds	GE1316	GTCTTCAACGA UGGTCCC	CTTCAAGGCT GACTGG	GTCTTCAACGAGGTCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTCACTCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CACTCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TGTGCTGCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TCACTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG TGGGCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTCTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG CTGCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG GGCAGGCTCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG ACTTCACTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG ACCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCTCACTGAGGCTCTCTGAGGCTCTCAACCTCTCAACCTCTGAGGCTGAGG	807

FIG. 5FFFFF

[illegible]

FIG. 5HHHHH

[illegible]

FIG. 51111

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
IGF1ul4	IGF1	T	G	-	-	noncoding	GE688	TCATAGCCTAG AAATGATCC TAT	AGGTTTGCAAT CAATTTGTTT	TCATAGCCTAGAAATGATCCCTATCTGAGATCAAGATTCTCTTCTAGAACCAATGAATTATCC AGCATTCAGATCTCTTCTAGTCACTTCTAGAACCTTTTGGTTTAAAGTACCAGGCTTGATTTATTC ATGCAAA[1/g]TCTATATTCTATCTCTGGAAGCTCTATATGAAAAACAAGAAATCACTCTTC AGTTTCTCCACTGGGTCACCTCAAGGATCAGAGGCCAGGAAAAAAGAGCTCCCTGGG TCTCTGATATATGCAAAAGAGGCCCTCTTGGAGCCAGCTCTAGTGGAGCCAGCACTCTCTCAGTCAACGAG TATTAACTCTCAGTCCAACTATTGTAATGAGCCCTCAGCATGCTTAGCAATGTTCTCAA TCACTATGAGCAGATGAAAGAACTATACTATTTTCCCTCTGCTGTTTCCAGACATA CAGGTTCTGAGGAAAGATCTGACTCTCTCTCCAGATGGCACTCTTTTATTTCTTCTC CCAGTGTACCTTTAAATTTCTCTCTCAACAACTTATAGGCACTCTCTCTGAGACT TAACTGTTCTCTGATGATGATGATACTCTCAAGATGCTATGATCTTCTCTCTCAG TCACTATCAGCAGCAAAATCCCAAGGAGAGCTGAAAGATGCACTGCCAATATA TCTTTTAACTTTTCCAACTTAACTCTCTCACTGATTTATATAAATGAAAAATAACT CATATACTCACTCTCTCTTTTAAAGATTAAGATTAAGTAAACAAATTTGATGCAACCC CT
IGF1ul5	IGF1	T	G	-	-	noncoding	GE688	TCATAGCCTAG AAATGATCC TAT	AGGTTTGCAAT CAATTTGTTT	TCATAGCCTAGAAATGATCCCTATCTGAGATCAAGATTCTCTTCTAGAACCAATGAATTATCC AGCATTCAGATCTCTTCTAGTCACTTCTAGAACCTTTTGGTTTAAAGTACCAGGCTTGATTTATTC ATGCAAA[1/g]TCTATATTCTATCTCTGGAAGCTCTATATGAAAAACAAGAAATCACTCTTC AGTTTCTCCACTGGGTCACCTCAAGGATCAGAGGCCAGGAAAAAAGAGCTCCCTGGG TCTCTGATATATGCAAAAGAGGCCCTCTTGGAGCCAGCTCTAGTGGAGCCAGCACTCTCTCAGTCAACGAG TATTAACTCTCAGTCCAACTATTGTAATGAGCCCTCAGCATGCTTAGCAATGTTCTCAA TCACTATGAGCAGATGAAAGAACTATACTATTTTCCCTCTGCTGTTTCCAGACATA CAGGTTCTGAGGAAAGATCTGACTCTCTCTCCAGATGGCACTCTTTTATTTCTTCTC CCAGTGTACCTTTAAATTTCTCTCTCAACAACTTATAGGCACTCTCTCTGAGACT TAACTGTTCTCTGATGATGATGATACTCTCAAGATGCTATGATCTTCTCTCTCAG TCACTATCAGCAGCAAAATCCCAAGGAGAGCTGAAAGATGCACTGCCAATATA TCTTTTAACTTTTCCAACTTAACTCTCTCACTGATTTATATAAATGAAAAATAACT CATATACTCACTCTCTCTTTTAAAGATTAAGATTAAGTAAACAAATTTGATGCAACCC CT
IGF1ul6	IGF1	T	C	-	-	noncoding	GE688	TCATAGCCTAG AAATGATCC TAT	AGGTTTGCAAT CAATTTGTTT	TCATAGCCTAGAAATGATCCCTATCTGAGATCAAGATTCTCTTCTAGAACCAATGAATTATCC AGCATTCAGATCTCTTCTAGTCACTTCTAGAACCTTTTGGTTTAAAGTACCAGGCTTGATTTATTC ATGCAAA[1/g]TCTATATTCTATCTCTGGAAGCTCTATATGAAAAACAAGAAATCACTCTTC AGTTTCTCCACTGGGTCACCTCAAGGATCAGAGGCCAGGAAAAAAGAGCTCCCTGGG TCTCTGATATATGCAAAAGAGGCCCTCTTGGAGCCAGCTCTAGTGGAGCCAGCACTCTCTCAGTCAACGAG TATTAACTCTCAGTCCAACTATTGTAATGAGCCCTCAGCATGCTTAGCAATGTTCTCAA TCACTATGAGCAGATGAAAGAACTATACTATTTTCCCTCTGCTGTTTCCAGACATA CAGGTTCTGAGGAAAGATCTGACTCTCTCTCCAGATGGCACTCTTTTATTTCTTCTC CCAGTGTACCTTTAAATTTCTCTCTCAACAACTTATAGGCACTCTCTCTGAGACT TAACTGTTCTCTGATGATGATGATACTCTCAAGATGCTATGATCTTCTCTCTCAG TCACTATCAGCAGCAAAATCCCAAGGAGAGCTGAAAGATGCACTGCCAATATA TCTTTTAACTTTTCCAACTTAACTCTCTCACTGATTTATATAAATGAAAAATAACT CATATACTCACTCTCTCTTTTAAAGATTAAGATTAAGTAAACAAATTTGATGCAACCC CT

FIG. 5JUUJ

122/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/ noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
IGF1u2 2	IGF1	T	A	-	-	noncoding	GE1191	TTTATAGGAGG TACATTTTGAG AAC	ACAACTACAA NATAGCACCAT	TTTATAGGAGGACATTTTGAAGACGcaagtagaggagtagcaggaacaggaactacagagatgt aggaaagccctctggaggagtagaggtagacatgcccgcaggatccttgcctctgcacaggtt acctgttaaaacttggaaacactacaaaataaagtttgatacaactttaaaagatgggctgttc ccccaatgaatcacacagtagaaacttccaaacttgcctttagggagtgatttgcaccttgcacaa atggtcctggagtgtagattgtgtgtgacttttatacaataatgttctatagaaaagaaaaa aaatataatataataatctctgcctcctcgaagcccaacaaatctcatcagagatgtgt atagatccagttgcaactaaatctctcctgaactctggcgtggagccatctcatcagcaact tgcttaagtggtttatgaattgttcttatttgaacttttctcaaacctcggctgtgtgt ttacagtgctgataaactgttagctataccacacacccctctcacaacttttataattgac cgaattggcctctcaaaagcagcagcagcgcgaagagcgaagcacaacaaatcttaaccacaagc ttccatctggcatcttaccaaa (t/a)ataagttggatgcatcttctttagacacaagct ttatcttccacacatcagcttacaaaaagataaagcaaatgttgcaacttggagccaaatc tttttaggcataatgtttaaacatagaaagtcttccaaactcaaaagagttccttcaaatgaga gttaatgcaacctaattagtaacttctctcttcttcttcttccatataagagcactatgaaa tttagcatcaatataacaggtatatacaaacagtagtanaactctgttttttagtataATGG TGCATTTTGTAGTTTGT
IGF1u2 3	IGF1	T	C	-	-	noncoding	GE1191	TTTATAGGAGG TACATTTTGAG AAC	ACAACTACAA NATAGCACCAT	TTTATAGGAGGACATTTTGAAGACGcaagtagaggagtagcaggaacaggaactacagagatgt aggaaagccctctggaggagtagaggtagacatgcccgcaggatccttgcctctgcacaggtt acctgttaaaacttggaaacactacaaaataaagtttgatacaactttaaaagatgggctgttc ccccaatgaatcacacagtagaaacttccaaacttgcctttagggagtgatttgcaccttgcacaa atggtcctggagtgtagattgtgtgtgacttttatacaataatgttctatagaaaagaaaaa aaatataatataataatctctgcctcctcgaagcccaacaaatctcatcagagatgtgt atagatccagttgcaactaaatctctcctgaactctggcgtggagccatctcatcagcaact tgcttaagtggtttatgaattgttctcttatttgcacttcttctacacactcgggctgtgtgt ttacagtgctgataaactgttagctataccacacacccctctcacaactcttatacttctg cgaattggcctctcaaaagcagcagcagcgcgaagtcgaagagcacaacaaatcttaaccacaag ttccatctggcatcttaccnaataaagttggatgcatcttattttagacacaaaagctttat tttccacatctgttacaanaagaaatagtaaaatagttgcaactttagggccaaatcattt taggcatactgtttaaacatagaaagtcttccaaactcaaaagagttcctcaaatgag (t/c)ga gttaatgcaacctaattagtaacttctctcttcttcttcttccatataagagcactatgaaa tttagcatcaatataacaggtatatacaaacagtagtanaactctgttttttagtataATGG TGCATTTTGTAGTTTGT
IGF1u2 4	IGF1	T	G	-	-	noncoding	GE683	TTTATATATAC TGAGGCTTAA AGT	AGATATACCAT TTTATATATAC ACTCT	TTTATATATACGAGCTTAAAGTaaacattctctcatttcttcttgcacaaaatgcaactgagt aaagtagaaaataaaaacagagctcaaaactccttcaagccacccattgaccccccactcacc aacctcatgcaaaagtcactctttaaactccttaactgatttggtagatattatcttctgac ccgctgctaaacacacacagcagggaggagctctgaacactcaagctgtctactacatcttttact gtgctgtgatacatgaaaatgtctatcaaaaatacaaaaccccttcaaaataacacgcagcttat attcagtttacaataaagggcccaataacacagtcagatcttttggtaaaagagtgaaagaaat tgagattggagattacatctgtatttgcctcatgtattttatcacactcaagggcaagagct gaataaactacagacactgaatatttccctgctactttagaacccagaaaaaatgact ggcaatctgtacatctgtctgttgaaagagcatatttattaaatattatctgtgtgtatt tgaactattatcaatcactatgagagaggaatacaactcactgaactctcaaaaatgtaa ctaattgaatcattat (t/g)cttacttactgtttaaagcatattttgaaaatgtatggcta GAGTGTCATATATAAATGTATATCT

FIG. 5MMMMMM

[illegible]

[illegible]

FIG. 500000

[illegible]

FIG. 5TTTTT

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
LIPC1 5	LIPC	G	C	G	A	cds	GE123	CTCCCGGTAA CCCTTACC	CGGCCATGAC TTCTATCTC	CTCCCGGTAAACCTTTACCCctctcttccatagggtgagatgcggggaacctttgttggagg gaagtccccccagcaatcgtcttctccagatagatgcgaattttgtggatgcttctacactttt accggggagcatatgggctggagctgggagctcaacacagctcataggacactatgacttctatctt caacggggctctctccagctc/g/c/tctgcaactcttagagctctacagacatatggccag caggcttcaatggTGAGAAATGAAGTATGGGCCG	295
LIPC1 6	LIPC	A	T	H	L	cds	GE174	TGAGAACCAAG TGATCTCTGA	ATCTGTGATCC TGCCCTTC	TGAGAACCAAGTATCTCTGAGcttggagctgtcttgggtgaagggtgataacgctcttttgc ctgtgttccagccatcccccagacatacaaaagctccacgagcgatcggtgcacttttccatct gactcttctgacggcgacgacagagc/a/t/lgggctaccctgctgtgagatgacatgaacagct tcagccagggtctgtctgagctgcaagaaggctgctgcaacagctgtgggtaccacgctccgc caggagccggagcagagacagagctcttctctgtaacgagccagctcccttcaaaagg tgagtgaggagctggggagccttcagAAGGCGCAGGATCGAGAT	369
LIPC1 7	LIPC	G	A	E	K	cds	GE175	AACCTGATTTG TCTGATTTTCT	TGGCACAAATG GGTGCTTA	AACTGATTTGTTGATTTTCTTctgtattcaaggggcaaggaaattgctagtaataaaacgta ttctcttttatacagctggtatgtgatactggagctgactcatgatcaagtccaagtgg/a laaacagtgagctgtgggcaatgtctggagacaggttccagaccatcatccatggagcaggg ggcgccactcagctctcttcaagagatcagagctcaagcagagagaaccagcaaggt gctgtgattcaactctcttattacgtctcatTAAGCACCCACTTGTGCCA	311
LIPC1 8	LIPC	G	T	A	S	cds	GE146	GCTGGAGAAGG AAGAGGGTA	TCACTCTCAGA GGAGGGAAA	GCTGGAGAAGGAAGGAGGTAGcggggagaggaagaaactaaggcgacctctctgtctccctc ctcaggtggagcggctgctagaaactggatctggcagatggcgcgctgaactcagcg ggccagccagtaagctggggctgtagtgaactgatacacttggccacagaccactacacatcgc cgctccacacccctcttgggcaaggaggtc/g/l/gggtctctctccgggtggctggaggta ccgacctggccagctctctcttccactctctctctctctctctCTTCTCTGAGATGA	320
LIPC1 9	LIPC	C	A	-	-	noncoding	GE123	CTCCCGGTAA CCCTTACC	CGGCCATGAC TTCTATCTC	CTCCCGGTAAACCTTTACCCctctcttctt/a/ccattaggctggatccggggacctttttt ggaggagttccccccagcaatcgtcttctccagatgatacccaattttgtgatgcaattctatc cttaccggggagcactgggctgagctgggagctcaaacagcccataggacactatgactctt atcccaacgggggtctctccagctggctggcactctctagagctctacagacatatggccag caggcttcaatggTGAGAAATGAAGTATGGGCCG	295
LIPC2	LIPC	T	A	V	E	cds	GE146	GCTGGAGAAGG AAGAGGGTA	TCACTCTCAGA GGAGGGAAA	GCTGGAGAAGGAAGGAGGTAGcggggagaggaaggaactaaggcgacctctctgtctccctc ctcaggtggagcggctgctagaaactggatctggcagatggcgcgctgaactcagcg ggccagccag/l/a/gaacgtgggctggtagctggatcaccttggccacagccactacacca tcgctctccgacacaccccttgggcaaggaggtcgggtctctctccgggtggctggaggta ccgacctggccagctctctcttccactctctctctctctctCTTCTCTGAGATGA	320
LIPC2 0	LIPC	A	T	I	F	cds	GE100	TGGGCAATCTT CCCTAACAA	ACCCCTGGATT CTTTGTGAC	TGGGCAATCTTCTCTAACMAgtatctataaggcatgtgtgtcttgggttcagaatatacca agaagcgttgagcccccgggggaacggggaagatggacacagctccctgtgttctccattctgt tggttttatgtcatctt/a/l/tccaatcaagtccttggaacagcctgaacagcgaatgaagag ctgtacttttctccagagatgggcatgaacttttcttttaaaacgtgtGTCAACAAAGAAATCCAG GGGT	264
LIPC2 1	LIPC	A	C	H	P	cds	GE153	CACATCGAAC GCAAAAGG	AGTGTGTGATT TATTAGGCATG G	CACATCGAACCGCAAAAGGctttctccagcggctctctctctctcccccaccccgctgctgc ttccaggaactgttccactctctcagaccatcttccactaatgttggtacagctggatgcac a/c/c/gctgtcaggtttgcccggcagttccatcggtagaacacacagattggggagaaatcacaggt aacCATGCTTAACTACTCACACT	220

FIG. 5VVVV

[illegible]

139/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
NCFB6	NCFB	T	C	-	-	noncoding	GE1186	TTTACAGAGCG GCTGACCTTG	GACACAGGCG TAACCG	TTTACAGAGCGCTGACGTTTGTCTACACATCTACAAGATGCGATGCGAGGCGCCAGT GAGGCGCCCTCCAGGCGAGCACTAATTCACATTAACIT/CJTGACCAAGTTGGGGATTATTT GTGGGTAACTGCGAGTGCAGTATGGAGTCTCTTGGGACAGTTAGAGCAATACCATTGATCTTA TAGTCACATAAGAACAAACAATAAAGAGAACAGACATGCTTAAGAGTGAAGAGAAAGGAGGGA GAAAGGAGGAGGTTGGATGGAGGACCTAGCTTGTAGTGAAGGTTCACTTGGATCTCATTC TGATTCAGTTTCAATTTGACTTCAGTCTTGTAGTGAAGTCACTTCTTCTGATCAGTTCC TTATGTGAATAAATAATGATAAATCACTAATGAACACAGACTCATGAGAGATAGAAGTGA ACACATTTAAAGAACATCACAAAGAGAACTATATGTGTCACATTTATATATGTGGGTGA GCTCTGAAGAGTCTGGACTAAGATGTTCCAGAGCCACAGGTTTTCACAAATGACGC TTGTGAATCATACAAGGCTCCCAAGTCCAGATCTTAGAGTGAACCTGCTGCTCTGGA AAGGGTTACCAGTCTGGGCTTCAAGCATGTCCTCCAGAGATCTCCCGTGCCTTCCAGA GGATCCAACTGTGAGCAGGACGCGACACATCAAGGACAGATGTCAGGGGAGAGGTGT AACTCTCCCAACCACTCTCTGGTACACATGGACACTTACCCTCTCTCTGAGGCTTAA GCTTCAGAGAACTCAAGAGCTGTGAATGATGTCCTCAAGCTCATATGAACTACTGGGCAAA ATTCCAGGGCTCTGCTCACTCTCTGGAGAGTCTGGATGGGTGACCACATCCATCTGCTG AGTCAGGCTCGGGTTACGCTGTGTC	1002
NCFB7	NCFB	C	T	-	-	noncoding	GE1186	TTTACAGAGCG GCTGACCTTG	GACACAGGCG TAACCG	TTTACAGAGCGCTGACGTTTGTCTACACATCTACAAGATGCGATGCGAGGCGCCAGT GAGGCGCCCTCCAGGCGAGCACTAATTCACATTAACIT/CJTGACCAAGTTGGGGATTATTTGG GTAACTGCGTGCAGTATGGAGTCTCTTGGGACAGTTAGAGCAATACCATTGATCTATAGT CACATGAACACAACTAAGAGAAAGACATGCTTAAGAGTGAAGAGAAAGGAGGAGAA GAGGAGGAGGTTGGATGGAGGACACTAGCTTAGTGAAGGTTCACTTGGATCTCATCTGCT TCAGTTTCAATTTGACTTCAGTCTTGTAGTGAAGTCACTTCTGATCAGTCTCTCTAT ATTTAAAGATCAATAATGATAAATCTCAATGACCTCAAGCTCATGAGAGATAGAAGTGAAC CTGAAGGTTGCTGGACTAAGTGGTCCAGGACCAAGGTTTTCACAAATGACGCTTGT TGAACTCAACAAGGCTCCCAAGTCACTAGATCTTAGAGTGAACCTGCTGCTGCTGAGG GGTACCACTCTGAGGCTCAAGACATGCTCCAGAGATCTTCTCTGCTCCAGGAGG GGATCAAACTGTGAGCAGGACGCGACACATCAAGGACAGATGTCAGGGAGAGGTGT AACTCTCCCAACCACTCTCTGGTACACATGGACACTTACCCTCTCTCTGAGGCTTAA GCTTCAGAGAACTCAAGAGCTGTGAATGATGTCCTCAAGCTCATATGAACTACTGGGCAAA ATTCCAGGGCTCTGCTCACTCTCTGGAGAGTCTGGATGGGTGACCACATCCATCTGCTG AGTCAGGCTCGGGTTACGCTGTGTC	1002

FIG. 5DDDDDD

[illegible]

[illegible]

[illegible]

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
PA12u3	PA12	T	G	P	P	cds	GE397	TCGTGTAAGTT CTATATCACCC ACA	GAAGTGTGTTCA GAAGACGAGAA AT	TCGTGTAAGTTCTATATCACCCACACATatagtaaaagtcacgtgattttttaaattagacattaaag tgttttctctcttcttcaataactatgctgaattttctgtttgtttgtttgtttgtttgtttgtttgtttg cagctggaagtgaaatacaactatgcaaaactcaacagtcacagtcgacgaagaacaaagagctgac agatgaagttgaggtatatacatcacccagcttcaaaattagaagagcattatgaactcagatccattc tgaagacatggatgtttcttctgaagttcccaagggccatgggtgaatgaagaggggac aggaatgacctgtttcttctgaagtttctgacaggaagacatggacatgaagggccacattttt tgaagacgctgctgacagggaggttatgacaggaagacatgaacatgaagggccacattttt tgcagatcatcctt/gt/ttcttttttattatgataaagataaccacatgcttattttt cgaagatttctccctcaacaaactaagcgtgcttctgcgaagaatttttgcataatttctgct gtgtgcttcgaagattgctatttcaaatgccaataatttagagatgtttctacatatttctgct CTTCTGACACACTTC
PA12u4	PA12	C	G	N	K	cds	GE397	TCGTGTAAGTT CTATATCACCC ACA	GAAGTGTGTTCA GAAGACGAGAA AT	TCGTGTAAGTTCTATATCACCCACACATatagtaaaagtcacgtgattttttaaattagacattaaag tgttttctctcttcttcaataactatgctgaattttctgtttgtttgtttgtttgtttgtttgtttgtttg cagctggaagtgaaatacaactatgcaaaactcaacagtcacagtcgacgaagaacaaagagctgac agatgaagttgaggtatatacatcacccagcttcaaaattagaagagcattatgaactcagatccattc tgaagacatggatgtttcttctgaagtttctgacaggaagacatggacatgaagggccacattttt tgaagacgctgctgacagggaggttatgacaggaagacatgaacatgaagggccacattttt tgcagatcatcctt/gt/ttcttttttattatgataaagataaccacatgcttattttt cgaagatttctccctcaacaaactaagcgtgcttctgcgaagaatttttgcataatttctgct gtgtgcttcgaagattgctatttcaaatgccaataatttagagatgtttctacatatttctgct CTTCTGACACACTTC
PA12u5	PA12	C	G	S	C	cds	GE397	TCGTGTAAGTT CTATATCACCC ACA	GAAGTGTGTTCA GAAGACGAGAA AT	TCGTGTAAGTTCTATATCACCCACACATatagtaaaagtcacgtgattttttaaattagacattaaag tgttttctctcttcttcaataactatgctgaattttctgtttgtttgtttgtttgtttgtttgtttgtttg cagctggaagtgaaatacaactatgcaaaactcaacagtcacagtcgacgaagaacaaagagctgac agatgaagttgaggtatatacatcacccagcttcaaaattagaagagcattatgaactcagatccattc tgaagacatggatgtttcttctgaagtttctgacaggaagacatggacatgaagggccacattttt tgaagacgctgctgacagggaggttatgacaggaagacatgaacatgaagggccacattttt tgcagatcatcctt/gt/ttcttttttattatgataaagataaccacatgcttattttt cgaagatttctccctcaacaaactaagcgtgcttctgcgaagaatttttgcataatttctgct gtgtgcttcgaagattgctatttcaaatgccaataatttagagatgtttctacatatttctgct CTTCTGACACACTTC
PA12u6	PA12	C	G	V	V	cds	GE312	GGGAAGACCAAT AATTCACCAT ATA	CTAAGTTCATG GATGGAATAAG ATA	GGGAAGACCAATTCACCAATTAagcattgctgtttgttatgtatttctatgacatttgcac ttttctgtctttaaaggaataatttgaactctgcagaaatttactctcagaacacaggcag tgaactctgaagatgycagangangactagaagaagaatttactctgttgcgtt/gt/gaagactca aaccgaagtgaaatccaagaataattttacttcttcttccagttagaagaactctgtatCTATC TTTTTCCATCATCAACTTAG

[illegible]

[illegible]

FIG. 5GGGGG

Poly ID	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
SLC6A4 u1	SLC6A4	G	C	K	N	cds	GE1124	TGGCTGGGCCA CCCTTA	TGGCTGGCCCC TAACAGG	CCGTTGGGCCACCCCTTAAactcgggctctgtctctccatcttaccactgcccaggattggcttatgcaatctgcatctgcttttctatgcttctctactacacacatcatgctctgggctgactactaccctcatctctccagagaccagctgctcggaccagctgcaa[g/c]aactcctggaaactgggaactgacacaattactctccagaggaacaatacacttggaccctccattccagctcccctgtggaagaattttacacgctgctgaagtcagtcagtaagtggggcttggcctgttgaaggccaggcca
SLC6A4 u2	SLC6A4	C	A	-	-	noncoding	GE1027	AGCATCAGTAA CCTGCACACTC	ATTTCGAAGCC CGACTGGT	AGCATCAGTAACTTCGCACACTCTctccctaggctctctcgaaggcaaggaggaacttgccttgcctctattgcaaatgaacaagggtctgaagagcaggttgc[ca/a]tggcaagtgggaagagaacaataggctggagcggcggaggagtttaaggAGCAGTCGGCTTCGCAAT
SLC6A4 u3	SLC6A4	A	C	-	-	noncoding	GE1027	AGCATCAGTAA CCTGCACACTC	ATTTCGAAGCC CGACTGGT	AGCATCAGTAACTTCGCACACTCTctccctaggctctctcgaaggcaaggaggaacttgccttgcctctattgcaaatgaacaagggtctgaagagcaggttgc[ca/a]tggcaagtgggaagagaacaataggctggagcggcggaggagtttaaggAGCAGTCGGCTTCGCAAT
SLC6A4 u4	SLC6A4	G	C	G	A	cds	GE1145	AAATCCAAAGCA CCACAGAT	ACTCCGGGTCA CAGCCAC	AAATCCAAAGCACCCAGATGcaatttggagctctggcagctggacatcagctgctcagcttgaagatgaagaacagcagatcgagacagcttgaatctcagaagaagctcatcagcgtctgaagctgaggaagcttggatcaggaagaggttctcagagaaggttgctccacccaggaggaagaagtggagctcgggcaaatgaataccaattgggtactcagcaggttccaagctctctgtctggg[g/a]agatgaacagcagctctatcccagacacacacccagctggctcagctcagaaggagagggagacacctggggcaagaaggtggaatttctctcagtaagtggctatgtggacctgggcaatctgttggcagctccctacatagtttaccagaatgaagggtgagtcagtcacagcgtgcgagtagtggtggtgtgacccaggctgtgacccaggct
SLC6A4 u5	SLC6A4	G	C	H	I	cds	GE1124	CCGTTGGGCCA CCCTTA	TGGCTGGCCCC TAACAGG	CCGTTGGGCCACCCCTTAAactcgggctctgtctctccatcttaccactgcccaggagattggcttatgcaatctgcatcttctatgcttctactacacacacatca[gc]gctctgggctcctatactcatctctctctcagggaccagctgctcctggaccagctgcccagcagctcgaagaactctggaaactgggaactgacaattactctcggaggaacaacatcacctggaccctccattccagctcccctgtcgaagaattttacacgctcagctgaagtcagtcagtaagggttggcctgttgaaggccaggcca
SLC6A4 u6	SLC6A4	C	T	F	F	cds	GE1124	CCGTTGGGCCA CCCTTA	TGGCTGGCCCC TAACAGG	CCGTTGGGCCACCCCTTAAactcgggctctgtctctccatcttaccactgcccaggagattggcttatgcaatctgcatcttctatgcttctactacacacacatca[gc]gctctgggctcctatactctctctcagggaccagctgctcctggaccagctcgaagaactctcgaagaactgggaactgccaactgactt[gc/t]tccgaggaacaacatcacctggaccctccattccagctcccctgtcgaagaattttacacgctcagctgaagtcagtcagtaagggttggcctgttgaaggccaggcca
SLC6A4 u7	SLC6A4	C	A	-	-	noncoding	GE1027	AGCATCAGTAA CCTGCACACTC	ATTTCGAAGCC CGACTGGT	AGCATCAGTAACTTCGCACACTCTctccctaggctctctcgaaggcaagggttgccttgcctctattgcaaatgaacaagggtctgaagagcaggttgc[ca/a]gaccttgccttgcctctatactgcaaaaatacaacagggggttgaagAGCAGTCGGCTTCGCAAT
SLC6A4 u8	SLC6A4	A	G	T	A	cds	GE1079	CTTAGACCCCT GATCTTGG	TCAAGCAAG CACTCAG	CTTAGACCCCTGATCTTGGAAactgtctcaggcggcccttggggttttccctccagagatgccttggg[a/g]cagagctggatgaaactgcatcagcagcttgccttgggatttgcatacttccacagctctcggcttccatctgagatgaagattgctcaggttggcccaagcagcaggttggagcctgggtctatgagcggccttgcctc[gc/tt]tga
TBXAS1 a15	TBXAS1	C	A	L	I	cds	GE282	CCTTCATCTGC AGCCATTT	TTGTCACGAT TATGACAGTT A	CCTTCATCTGCACCCATTTagggtagctcccgacagctgctcctaactacaccttggcttatccatctcatcagatggttccctcatcagcaacactcagacact[ca/a]tccatgctcattcaaacgcttatcgggaaactcgggacgatttgacaaccagggatcaaggtgtgctgcatctacagatgagaatcaggattttgaaactcactgctcttggTAAGTGTCCATAATTCGTGACGA

FIG. 5

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence
TBXAS1 la16	TBXAS 1	A	G	-	-	noncoding	GE332	CGAGATTGAAA TTTAAAGGAAA GAC	CCGAAACACA AGTGGTAACTG A	CGAGATTGAAATTTAGGAAAGACaaaatgctgtgagattgggtgtaacacga [a/g]tctctc 289 ccttggacagacccctccatcagatggccctgagttctcagcctcagggaagccctgcacctc tctgacatggtgatgcagagacgtgagatgatacccgacagcttccagtggtggtgagagccc ctccctgcctgaggtcccccctcctaccctaccctgcccagcctcaggttcaggggccct ctccatcAGTTTACCCTTGTGTCTTCGG
TBXAS1 la17	TBXAS 1	C	T	-	-	noncoding	GE332	CGAGATTGAAA TTTAAAGGAAA GAC	CCGAAACACA AGTGGTAACTG A	TGTCATTGAAATTTAGGAAAGACaaaatgctgtgagattgggtgtaacacgaacattctccct 289 tgcacagccctccat [c/t] agatggccctcagttctcagcctcagggaagccctgcacctc tctgacatggtgatgcagagacgtgagatgatacccgacagcttccagtggtggtgagagccc ctccctgcctgaggtcccccctcctaccctaccctgcccagcctcaggttcaggggccct ctccatcAGTTTACCCTTGTGTCTTCGG
TBXAS1 dl12	TBXAS 1	C	G	Q	E	cds	GE357	TGCTGTTCCLAA ATTGTTTACTG	TTTCACACCTG AAATATCAAAAT	TGCTGTTCAAATTTTACTGAAATgaatgtgaaatggaatttggctcaattcttcttcttctac 222 tatagtgctgtctcttcaggggtttttgggaagaag [c/g] aaatggagctcagaagaagctga tggccctgtgtggtggaagaagaacctcaacgctctctat [a/g] aagatatttctat [a/t] g acgatATTTTGATTTTACCTGTTCGA
TBXAS1 dl13	TBXAS 1	G	A	R	H	cds	GE326	TGGAAACCTAT TCTTTGCTCTT	TACAGCCATGA GCCACTGT	TGGAAACCTATTTCTTTGCTCTTtacttcagagagctcag [a/t] cttcaggttctcctaatagagcct 288 aaagctgagtgcaacttcaattctcagcttttgaatctgttttctcctcaggttctcctacat cagcatctcagaagctggagagatgaggtcagacatcccaagcttctctcttcaattggaaac ttgacattttcc [g/a] ccagtaagggtgtcttccattggcttccatcataataaata [a/t] gctga ggccaggCACAGTGGCTCTATGGCTCTA
TBXAS1 dl14	TBXAS 1	G	A	-	-	noncoding	GE355	CTTGAGCATC CTTGCTCTCA	GCTCTCAGCA GAGAACTGG	CTTGAGCATCTTGTCTCTCAga [c/g] cagggtggtctcagtcagacacagggctcagagggggag 347 gagcgggtgtcttgggcagccctgcacacagagcctcaggttcaagctgagccggcggcagc agacccgctctcagctaccttgccttccggggccggccacagagctgcctcgggtgctgacta gggtgtctgaggtcaagttgacatctcagctgtcagcaagttccggttccacagcttgccc tgagaccaggtgagggccctcgtcagagggcag [g/a] tcaaggggcagcgttggggggccca ccCAGTTCTCTGCTGAGAGC
TBXAS1 ul1	TBXAS 1	G	C	V	L	cds	GE355	CTTGAGCATC CTTGCTCTCA	GCTCTCAGCA GAGAACTGG	CTTGAGCATCTTGTCTCTCAga [c/g] cagggtggtctcagtcagacacagggctcagagggggag 347 gagcgggtgtcttgggcagccctgcacacagagcctcaggttcaagctgagccggcggcagc agacccgctctcagctaccttgccttccggggccggccacagagctgcctcgggtgctgacta gggtgtctgaggtcaagttgacatctcagctgtcagcaagttccggttccacagcttgccc ggcctgagaccaggtgagggccctcgtcagagggcaggtcagagggcagcgttggggggccca ccCAGTTCTCTGCTGAGAGC
TBXAS1 ul10	TBXAS 1	A	C	-	-	noncoding	GE249	ATGCACTGTA TTGCCACCA	GAGAGTTTGA TTTCTCATGTC TTA	ATGCACTGTAATGCACTGTAagttggttcttggctcctcagtcctgacccctctggtt [a/c] 219 cttccacacagcggcgggttttgggttcaagtcagtcagtcagcagcgttctggttttctacgtac aaaga [c/g] ggaagagcaggtcagaggtgacctga [c/t] gctgtcttctcagtcctgaaagcgtgaacaggt AAGACATGAGAAATGCAAACTCTC
TBXAS1 ul11	TBXAS 1	C	T	T	M	cds	GE355	CTTGAGCATC CTTGCTCTCA	GCTCTCAGCA GAGAACTGG	CTTGAGCATCTTGTCTCTCAga [c/g] cagggtggtctcagtcagacacagggctcagagggggag 347 gagcgggtgtcttgggcagccctgcacacagagcctcaggttcaagctgagccggcggcagc agcagggcctc [c/t] gtaacctccttggggccggccacagcagcgttccgtcgggtgctga tctagggtgcttgggttcaagttgacatctcagctgtcagcaagttccggttccagcct ggcctgagaccaggtgagggccctcgtcagagcaggtcagagggcagcgttggggggccca ccCAGTTCTCTGCTGAGAGC

FIG. 5JJJJJ

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Assay Sequence	
THBDu1	THBD	T	G	C	G	cds	GE409	TGCGCTTACAG CTAATGTGC	GCCAGCTAAGG TGCTTTGGT	TCGGCTTACAGCTAATGTGCaccgctccgcccggagcgggtccaggggcactgggcccaggagcgcg ccggcgcttggagactgcagctggagaaacggcggtctgagacgcgtgcgaatgcgactccctggg ggctcccgcctgcagatgccacgactctgcagccctgcagcgagacgggcgtctctctccacgcgat ccgacgcagctctctgcgaacgactctgcgagcactcttgcgttcccaaccgccacgacgcgggc tctactcgtgcactgcgagacggctaccggctggcgccgcacacacacggctggcgagagcgt ggatgactgcatactggagcccacttcggtctgcgagcgtgtgttcaaacacaggggtggcttcg agtgcacactgctacctaaactaactagttggacggcggtt/gtggtagagacgggtggaccc gtgcttcagagccaactgcagctacggttcgagccctgaaccaactgcatactctctgcgtct ggcgagggtcttcggcccaattccnccagcgccgcacaggttgcacagatgttttgcacaacagact gcctatccagcgactgcgacccaacacacggctagctgtgagtgccctgagagctacatcct ggcgacgggtttcatctgcagcgaacttcgcgggtgcgaacacggcggtctctgcgcgggggt ggcaaacctcccgtaccttcgagttgcattgcggggcccgactcggcccttgcgcgcacacatt ggcacgactgtgacttcggcaaggctggacggttggcgacagcagctcttcggagcccccgcacg cccgacggcggttccaccttgactctccggccttggggctcgtgcattcggggttgctctcag gcatctcatcgcgagcctgtgcttgttgtggcgcttttggcgctctctctgcacactgcgcgaag agcgccggcgccggggcgccgaagtggatgcagatgctggcgccccctt	1164
THBDu2	THBD	C	A	P	T	cds	GE409	TCGGCTTACAG CTAATGTGC	GCCAGCTAAGG TGCTTTGGT	TCGGCTTACAGCTAATGTGCaccgctccgcccggagcgggtccaggggcactgggcccaggagcgcg ccggcgcttggagactgcagctggagaaacggcggtctgcgagacgctgcgaatgcgactccctgg ggctcccgcctgcagtgcccgacgttcggcgctggcgccctgcagcgagcggcggtctctgcacgcgat cccgacgcagctcctgcgaacgactctgcgagcactcttgcgttcccaaccgccacgacgcggcc tctactcgtgcactgcgagacggctaccggctggcgccgcacacacacggctggcgagagcgt ggatgactgcatactggagcccacttcggtctgcgagcgtgtgttcaaacacaggggtggcttcg agtgcacactgctacctaaactaactcgaactggtagcgcggtgtgtggagacggcgtgcgacccgtgc ttcagagccaactgcagtagaccagtcag[c/a]ccctgacccaactagtagtacctctgcgtct ggcgagggtcttcggcccaattccnccagcgccgcacaggttcgagatgttttgcacaacgact gcctgtccagcgactgcgacccaacacacggctagctgtgagtgccctgcgctacatcct ggacgaggtttcatctgcagcgaactgcagagtagtcgcaaacacggcggtcttcgctccgggggt ggcaaacctcccgtaccttcgagttgcattgcggggcccgactcggcccttgcgcgcacatt ggcacgactgtgacttcggcaaggctggacgggtggcgacagcgggtcttcggagcccccgcacg cccgacggcggttccaccttgactctccggccttggggctcgtgcattcggggttgctctcag gcatctcatcgcgagcctgtgcttgttgtggcgcttttggcgctctctctgcacactgcgcgaag agcgccggcgccggggcgccgaagtggatgcagatgctggcgccccctt	1164

FIG. 5LLLLLL

[illegible]

FIG. 5

175/178

Poly Id	Gene	ref Nr	alt Nr	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence
THPO1	THPO	C	7	-	-	noncoding	GE345	TCACTCTGCTG GCTACTCTTA	CCTTCTTCCCT CAGGTCTCTTA	TCACTCTGCTGCTACTCTTAAGggccccccgcttttagtgcccccttgaggcagtgccgtt tctcttccatctctcttcaggaggagccagccagagcattcttgaggcagtgaccctt ctgctggaggaglgatggcagcagggagcaactggagccacttgcctctccctccctggg gcagcttctggagacaggtccgtctctctctctggggccctgcagagcctcttggaacccaggtaa gtcccagtcagggaactgtgtggaactgttcttctgactcagtcctc/ICTAGACAGCTG AGGAAAGAGG
THPO1	THPO	T	A	S	T	cds	GE416	GGCATCTCTGTC TTTCTTACTTA GAC	AGGAATCTTTC TCCAGTTGTC C	GGCATCTCTCTTCTTACTTACTAGCaaggaggccctgagatctggcccttggtgttgccctcagg accatctctctgctctcagcttctctccacagaggagccagccagctcacaaggatcccaatgccat cttctgagcttccaacactgtctcaggagggtgcttctctgagcttgtagagggtccca cccttgctcagggcccccacccacagctgtccccagcagaaactctctagctctcaca ctgaacagctcccaaacaggacttctgattgttgagacaaacttcactgctcagccagaaac gaactctggctctgagcttctgagctggcagcaglg/a/gattcagagccagattcttggtctgtg aaccacccctcagggtccctggaccaaatcccaggataactgaacagatcacgaactcttga tggactctgaggacttctctggacctccacagcagcccttaggagagcccgagacatttccctcag gaactcagacacaggttctctgcccccaactccagcctggatattctcttcccacccat cctctactggagatgacgtcttctcttccaccccttgcaccccttccagcctctcttaaacat ccacccctgcttctgagacttctgctcagcagcccccctaccagcctctcttaaacat cctacacccatcccagatctgtctcaggagggttaagggtctcagacactgctgacatcagca ttgtctctgtgacagctcccttctctcaggggccctctggGAGACACTGGACAGATTCTCT
THPO2	THPO	A	G	Q	R	cds	GE265	TGGAGGACTAG CCTGCTTATTA	AGAATCCATGG GAAGCAGTG	TGGAGGACTAGCCTGCTTATTAAGgctacacatagctctctctattcagctcccttccccccac caatcttctcagagcc/a/gtgccccagagggttcccttctgctacactgtctctg cctgctgagctttagcttggagatggaaacccagatggtagaaggccatccctaaacctt ggcttccctaaagctgttctcagttctcCACTGCTTCCCATGGATTCT
THPO3	THPO	G	A	G	R	cds	GE416	GGCATCTCTGTC TTTCTTACTTA GAC	AGGAATCTTTC TCCAGTTGTC C	GGCATCTCTCTTCTTACTTACTAGCaaggaggccctgagatctggcccttggtgttgccctcagg accatctctgctctcagcttctctccacagaggagccagccagctcacaaggatcccaatgccat cttctgagcttccaacactgtctcaggagggtgcttctctgagcttgtagagggtccca cccttgctcagggggggcccccacccacagctgtccccagcagaaactctctagctctcaca ctgaacagctcccaaacaggacttctgattgttgagacaaacttcactgctcagccagaaac gaactctggctctgagcttctgagctggcagcaglg/a/gattcagagccagattcttggtctgtg aaccacccctcagggtccctggaccaaatcccaggataactgaacagatcacgaactcttga tggactctgaggacttctctggacctccacagcagcccttaggagagcccgagacatttccctcag gaactcagacacaggttctctgcccccaactccagcctggatattctcttcccacccat cctctactggagatgacgtcttctcttccaccccttgcaccccttccagcctctcttaaacat ccacccctgcttctgagacttctgctcagcagcccccctaccagcctctcttaaacat cctacacccatcccagatctgtctcaggagggttaagggtctcagacactgctgacatcagca ttgtctctgtgacagctcccttctctcaggggccctctggGAGACACTGGACAGATTCTCT

FIG. 5NNNNNNN

176/178

Poly Id	Gene	ref NT	alt NT	ref AA	alt AA	coding/noncoding	Assay #	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Assay Sequence	
THP04	THPO4	A	T	K	M	cds	GE16	GGCATCTCTGTC TTTCTTACTTGA GAC	AGGAATCTTG TCCAGTGTCT C	GGCATCTCTGTCCTTACTTACAGaaggagcctgagatctggccttggtgtttggcctcagg accatctctgctctcaagctctctccacagggagagacacagctcccaagatcccaatgccat ctctctgagcttccacacactctccaggaa/fa/ttgggtggttctctgagatcttgaaggggg tccacctctgctgagggggcccccacacagctgtcccagcagaaactctctctagctct cacactgaacgagctcccaaacaggactctggatgttggagacaactcactctctcagcca gaactactggtctgggctctgaagtggcagcagagattcagacgaagatctctggtctgtg aaccaacctccaggctctggacaaatccccggataactgaacagatacacgaactctgaa tggaaactctggagctctctggacccctcagcagagacctaggagccccggacattctctcag gaactcagacacaggtctctcgcacccaactctcagcttggaattctctctccccaaacct ctctctactggacagatctgctctctctctccacaccttgccacaccttggtccagct ccacacctgcttctgaacctctgctctcacaagcagctctcagacacctctctaaacacat cttaacacctcccaaatctgctcagagggtaaggttctcagacacctccgacatcagca ctgtctgctgacagctctctctctcagggcgcctctggagacactccgacatggtctct tctgtctgctgacagctctctctctcagggcgcctctggagacactccgacatggtctct GGGCT	844
THU1	TH	G	A	S	S	cds	GE1125	GGTCCCGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCGGGTCTCAGCaggtgagagagcattgggtgctctgtctccacagctccccgggttc attgggcgcagcagagcctctcagagaccccgcaaggagcggcgaggtggcagcagc ggcgtctgagctccctcagagcccgaggaccccttgaggtgtggtctttgagagagagag ggaaggcctg/a/tgctaaacctgctctctcccgaggccacacagcctcggcgtctcccg agctgtgaagggtttgaggtgagctgggtgctctctcctggggcaagtTCACCTGTGGGT GGGCT	331
THU2	TH	G	A	V	M	cds	GE1125	GGTCCCGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCGGGTCTCAGCaggtgagagagcattgggtgctctgtctccacagctccccgggttc attgggcgcagcagagcctctcagagaccccgcaaggagcggcgaggtggcagcagc ggcgtctgagctccctcagagcccgaggaccccttgaggtgtggtctttgagagagagag ggaaggcctg/a/tgctaaacctgctctctcccgaggccacacagcctcggcgtctcccg agctgtgaagggtttgaggtgagctgggtgctctctcctggggcaagtTCACCTGTGGGT GGGCT	331
THU3	TH	T	G	F	C	cds	GE1125	GGTCCCGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCGGGTCTCAGCaggtgagagagcattgggtgctctgtctccacagctccccgggttc attgggcgcagcagagcctctcagagaccccgcaaggagcggcgaggtggcagcagc ggcgtctgagctccctcagagcccgaggaccccttgaggtgtggtctttgagagagagag ggaaggcctg/a/tgctaaacctgctctctcccgaggccacacagcctcggcgtctcccg agctgtgaagggtttgaggtgagctgggtgctctctcctggggcaagtTCACCTGTGGGT GGGCT	331
THU4	TH	G	A	K	K	cds	GE1020	CTGCCCCGAGG AAGGAG	CTGGGCACAC CTTCAG	CTGCCCCGAGGAGAGgtctacacacgctgaa/g/a/ggctctacgcacagcagcctgcgg ggagcactggaggcctttgctttgtgagagctcagcggctcagggagagacaatacccccc agctggaggagcgtctcccgcttctcTGAAGGGTGTCCCAAG	170
THU5	TH	T	C	A	A	cds	GE1125	GGTCCCGGGT CTCAGC	AGCCCCACCA CAGGTGA	GGTCCCGGGTCTCAGCaggtgagagagcattgggtgctctgtctccacagctccccgggttc attgggcgcagcagagcctctcagagaccccgcaaggagcggcgaggtggcagcagc ggcgtctgagctccctcagagcccgaggaccccttgaggtgtggtctttgagagagagag ggaaggcctg/a/tgctaaacctgctctctcccgaggccacacagcctcggcgtctcccg agctgtgaagggtttgaggtgagctgggtgctctctcctggggcaagtTCACCTGTGGGT GGGCT	331
THU6	TH	G	T	A	A	cds	GE972	TACGCCGAGG ACTGCT	GGGTACTGCG AGTACT	TACGCCGAGGAGTCTGccacagctgctggggcagctgccccatgctggcagcagccttcg c/g/t/cAGTTCTGCGAGGTACAGC	89

FIG. 50000000

[illegible]

FIG. 5PPPPPP

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 October 2000 (05.10.2000)

PCT

(10) International Publication Number
WO 00/58519 A3

(51) International Patent Classification⁷: C12Q 1/68

SKLAR, Pamela [US/US]; 68 Fuller Street, Brookline, MA 02446 (US).

(21) International Application Number: PCT/US00/08440

(22) International Filing Date: 30 March 2000 (30.03.2000)

(74) Agents: HOGLE, Doreen, M. et al.; Hamilton, Brook, Smith & Reynolds, P.C., Two Militia Drive, Lexington, MA 02421 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/127,248 31 March 1999 (31.03.1999) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(71) Applicants (*for all designated States except US*): WHITE-HEAD INSTITUTE FOR BIOMEDICAL RESEARCH [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US). AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): ALTSHULER, David [US/US]; 69 Mason Terrace, Brookline, MA 02146 (US). CARGILL, Michele [US/US]; One Warwick Park #3, Cambridge, MA 02140 (US). DALEY, George, Q. [US/US]; 50 Young Road, Weston, MA 02493 (US). IRELAND, James, S. [US/US]; One Warwick Park, #3, Cambridge, MA 02140 (US). LANDER, Eric, S. [US/US]; 151 Bishop Allen Drive, Cambridge, MA 02139 (US). LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US). PATIL, Nila [US/US]; 780 West California Way, Woodside, CA 94602 (US).

Published:

— with international search report

(88) Date of publication of the international search report:
23 August 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 00/58519 A3

(54) Title: CHARACTERIZATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN CODING REGIONS OF HUMAN GENES

(57) Abstract: The invention provides nucleic acid segments of the human genome, particularly nucleic acid segments from the coding region of a gene, including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking or containing these sites are also provided. The nucleic acids, primers and probes are used in applications such as phenotype correlations, forensics, paternity testing, medicine and genetic analysis.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/08440

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 96 20279 A (RIBOZYME PHARM INC ;WARNER LAMBERT CO (US)) 4 July 1996 (1996-07-04) p. 15, sequence "013a"	1
X	US 5 705 388 A (COUTURE L. ET AL.) 6 January 1998 (1998-01-06) SEQ ID 622,623	1

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "Z" document member of the same patent family

Date of the actual completion of the international search

10 November 2000

Date of mailing of the international search report

22.02.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Luzzatto, E

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/08440

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>DATABASE MEDLINE [Online] American medical association Accession Number 92172843, 3 March 1992 (1992-03-03) XP002152537 abstract & SUMI-ICHINOSE C. ET AL.: "Molecular cloning of genomic DNA and chromosomal assignment of the gene for human aromatic L-amino acid decarboxylase, the enzyme for catecholamine and serotonin biosynthesis" BIOCHEMISTRY, vol. 31, no. 8, 1992, pages 2229-2238, US</p>	1,11
A	<p>--- CHEE M ET AL: "ACCESSING GENETIC INFORMATION WITH HIGH-DENSITY DNA ARRAYS" SCIENCE,US,AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, vol. 274, 25 October 1996 (1996-10-25), pages 610-614, XP002022508 ISSN: 0036-8075 cited in the application the whole document</p>	1-12
A	<p>--- WANG D G ET AL: "Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome" SCIENCE,US,AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, vol. 280, 1998, pages 1077-1082, XP002089398 ISSN: 0036-8075 cited in the application the whole document</p>	1-12
A	<p>--- OEFNER P J ET AL: "COMPARATIVE DNA SEQUENCING BY DENATURING HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY" AMERICAN JOURNAL OF HUMAN GENETICS,UNIVERSITY OF CHICAGO PRESS, CHICAGO,,US, vol. 57, October 1995 (1995-10), page A266 XP002918646 ISSN: 0002-9297 cited in the application the whole document</p>	1-12
A	<p>--- WO 98 20165 A (WHITEHEAD BIOMEDICAL INST ;HUDSON THOMAS (US); LANDER ERIC S (US);) 14 May 1998 (1998-05-14) page 4, line 1 -page 27, line 23 --- -/--</p>	1-12

INTERNATIONAL SEARCH REPORT

Internal Application No
PCT/US 00/08440

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>DATABASE GENE BANK [Online] accession number M84597, 18 April 1992 (1992-04-18) ICHINOSE H. ET AL.: "Molecular cloning of genomic DNA and chromosomal assignment of the gene for human aromatic L-amino acid decarboxylase" XP002152538 abstract</p>	
A	<p>SYVANEN A -CH ET AL.: "IDENTIFICATION OF INDIVIDUALS BY ANALYSIS OF BIALLELIC DNA MARKERS, USING PCR AND SOLID-PHASE MINISEQUENCING" AMERICAN JOURNAL OF HUMAN GENETICS, US, UNIVERSITY OF CHICAGO PRESS, CHICAGO,, vol. 52, no. 1, 1993, pages 46-59, XP002050638 ISSN: 0002-9297 the whole document</p>	11,12
A	<p>UNDERHILL P.A. ET AL.: "Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography" GENOME RESEARCH, vol. 7, - 1997 pages 996-1005, XP000942730 US abstract</p>	1-12

INTERNATIONAL SEARCH REPORT

In. ational application No.
PCT/US 00/08440

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-12 all partly

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-12 (all partly)

Nucleic acid molecules comprising the polymorphic sites present in the sequences SEQ ID 1927 and 1928 given on fig. 5A (which correspond to the sequences designated AADCd4 and AADCd5 in fig. 5A), and methods based thereon.

2. Claims: 1-12 (all partly)

Nucleic acid molecules comprising the polymorphic site(s) present in each of the sequences given on figs. 5A (except the sequences AADCd4 and AADCd5), 5B-5QQQQQQ and methods based thereon.

The sequences differing only at one or more polymorphic sites are considered to belong to the same invention. The different inventions are summarised as indicated in the following table (see annex), where col. 1 and 3 show the SEQ ID number and col. 2 and 4 the length of the sequence. The total number of inventions is 396.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/08440

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9620279 A	04-07-1996	US 5705388 A AU 4419596 A CA 2208502 A EP 0800580 A JP 10511269 T	06-01-1998 19-07-1996 04-07-1996 15-10-1997 04-11-1998
US 5705388 A	06-01-1998	AU 4419596 A CA 2208502 A EP 0800580 A JP 10511269 T WO 9620279 A	19-07-1996 04-07-1996 15-10-1997 04-11-1998 04-07-1996
WO 9820165 A	14-05-1998	EP 0941366 A	15-09-1999